



Published in final edited form as:

Nat Genet. 2009 November ; 41(11): 1191–1198. doi:10.1038/ng.466.

Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium

A full list of authors and affiliations appears at the end of the article.

Abstract

Users may view, print, copy, download and text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence should be addressed to Santhi K. Ganesh, M.D. (ganeshs@mail.nih.gov) or Christopher J. O'Donnell, M.D., M.P.H. (odonnelle@mail.nih.gov).

[^]Authors indicated by a caret (^) contributed equally as senior authors

Authors indicated by an asterisk () contributed equally as junior authors

Author Contributions

CHARGE:

AGES: Study design and phenotype collection, TBH, VG, LJJ; Genotyping, ; data analysis, AVS, MAN, TA; manuscript preparation, MAN, NS; manuscript revisions, TBH, VG, LJJ, AVS, ABS, DGH, MAN

ARIC: Study design and phenotype collection, EB, JC, SF, AC; Genotyping, EB, AC; data analysis, SKG, AK, AC, AG, GBE; manuscript preparation, SKG, SF; manuscript revisions, SKG, AC, EB, JC, AK, SF.

CHS: Study design and phenotype collection, BMP, JIR, MC, NAZ, NLG; Genotyping, JIR; data analysis, NLG, TL; manuscript preparation, NAZ; manuscript revisions BMP, JCB, KR, JLR, MC, NAZ, NLG, TL.

FHS: Study design and phenotype collection, MHC, QY, CSF, DL, LAC, CJO, JPL; Genotyping, ; data analysis, MHC, QY, JPL; manuscript preparation, MHC, CJO, JPL; manuscript revisions, MHC, QY, CSF, DL, LAC, CJO, JPL.

Rotterdam: Study design and phenotype collection, FJAR, AH, AGU, BAO, CMD, JCMW; Genotyping, AGU; data analysis, FJAR, JFF, AD, GCV; manuscript preparation, FJAR, JFF, CMD; manuscript revisions FJAR, JFF, AD, GCV, AH, AGU, BAO, CMD, JCMW

InCHIANTI: Study design and phenotype collection, LF, JMG, SB; Genotyping, ; data analysis, MAN, TT; manuscript preparation, MAN; Manuscript revisions, LF, JMG, SB, KVP, ABS, DGH, MAN, TT

HaemGen:

Genotyping, AR, JJZ, JMvG

TwinsUK: Study design and phenotype collection, TDS, SLT, PD; data analysis, NS; manuscript preparation NS, manuscript revisions, NS

UKBS1: Study design and phenotype collection, WHO, JGS; data analysis, NS; manuscript preparation NS, AR, manuscript revisions, NS, AR

KORA: Study design and phenotype collection, CM, CG; data analysis, CG, BK; manuscript preparation CG, manuscript revisions, CG

SHIP: Study design and phenotype collection, MN, AG; data analysis, AT

Gene expression: AR, JJZ, JMvG

Competing Interest Statement: Aravinda Chakravarti is a paid consultant with Affymetrix in accordance with the policies of Johns Hopkins.

URLs:

AGES, <http://www.hjarta.is/english/ages>;

ARIC, <http://www.csc.unc.edu/aric/>;

Cardiovascular Health Study, <http://www.chs-nhlbi.org/>;

Framingham Heart Study, <http://www.framinghamheartstudy.org/about/index.html>;

Rotterdam Study, <http://www.epib.nl/ergo.htm>;

BIMBAM, <http://stephenslab.uchicago.edu/software.html>;

EIGENSTRAT, <http://genepath.med.harvard.edu/~reich/Software.htm>;

GenABEL and ProbABEL, <http://mga.bionet.nsc.ru/~yurii/ABEL/>;

HapMap, <http://hapmap.org/>;

MACH v1.0.15/16 (<http://www.sph.umich.edu/csg/abecasis/MaCH/index.html>);

PLINK, <http://pngu.mgh.harvard.edu/purcell/plink>.

METAL, <http://www.sph.umich.edu/csg/abecasis/Metal/index.html>

SNAP, <http://www.broadinstitute.org/mpg/snap>

Illumina BeadChip Probe Annotation <http://www.compbio.group.cam.ac.uk/Resources/Annotation/>.

Erythrocyte measures are heritable and have important health implications, yet their genetic determinants are largely unknown. We performed genome-wide association analyses in 24,167 European-ancestry individuals for six erythrocyte traits and identified associations at 23 loci (P values 5×10^{-8} to 1×10^{-57}). Replication testing in an independent set of 9,456 European-ancestry individuals showed strong evidence of association in all 23 loci in meta-analysis of the discovery and replication cohorts. Our findings include previously identified loci (*HBS1L/MYB*, *HFE*, *TMPRSS6*, *TFR2*, *SPTA1*) and novel associations (*EPO*, *TFRC*, *SH2B3*, and 15 other loci). This study has identified novel determinants of erythrocyte traits, offering insight into common variants underlying variation in erythrocyte measures.

Introduction

Red blood cell disorders such as anemia and erythrocytosis are broadly associated with multiple comorbid conditions including hypertension and other cardiovascular diseases, yet the genetic determinants of erythrocyte traits in the general population are poorly defined. Erythrocytes, which comprise approximately 40% - 50% of blood volume, are a key component for the transport of oxygen and carbon dioxide for cellular respiration. In clinical practice, measures of erythrocyte quantity, size and composition are routinely tested to diagnose and monitor hematologic diseases as well as the overall health of patients. Variation in erythrocyte measures even within normal ranges are related to other non-hematologic diseases and mortality¹⁻³.

Erythrocyte production and quality are under various environmental and genetic influences. While environmental exposures, dietary intake of vitamins and iron, and the anemia of chronic disease contribute substantially to abnormalities of erythrocyte measures, the heritability of erythrocyte traits ranges from 40% - 90%⁴⁻⁶. Disorders of hemoglobin production and hemoglobinopathies are some of the most common genetic diseases in the world, owing to natural selection. Some known low-frequency Mendelian variants also lead to inter-individual variability in erythrocyte traits in the general population^{7, 8}. Candidate gene studies have identified a few non-hemoglobin loci, including *EPOR* and *HBS1L*, related to variation in erythrocyte traits⁸⁻¹⁰. Early genome-wide association and linkage studies of erythrocyte measures, which have identified a few associations, such as at chromosome 6q23, lacked statistical power for association and genetic resolution for testing competing hypotheses^{6, 11-13}.

To investigate genetic determinants of erythrocyte traits in the general population, we carried out genome-wide association studies and meta-analysis within multiple community-based cohorts comprising the CHARGE consortium¹⁴, followed by replication in independent samples. We identified 23 genetic loci associated with these erythrocyte traits. We further extend these findings to investigate possible links between these traits and vascular diseases, reporting associations of a few of the 23 loci identified with blood pressure and hypertension.

Results

CHARGE Consortium study samples

The total sample size for the individual cohort genome-wide association analysis and the CHARGE meta-analysis was 24,167 (the Age, Gene/Environment Susceptibility Reykjavik Study (AGES), N=3,205; the Atherosclerosis Risk in Communities Study (ARIC), N=7,803; the Cardiovascular Health Study (CHS), N=3,256; the Framingham Heart Study (FHS), N=3,359; and the Rotterdam Study (RS), N=5,523). We also included the Invecchiare in Chianti Study (InCHIANTI, N=1,021), an Italian cohort study, in these analyses. Characteristics of the study participants, including age, sex and trait summaries, are presented in Table 1.

Meta-analysis of genome-wide association studies for six erythrocyte traits in the CHARGE Consortium

We studied six erythrocyte traits: hemoglobin concentration (Hgb); hematocrit (Hct); mean corpuscular volume (MCV); mean corpuscular hemoglobin (MCH); mean corpuscular hemoglobin concentration (MCHC); and red blood cell count (RBC), as defined in Supplementary Table 1. When cohort results were combined, 831 SNP associations at 23 independent loci ($r^2 < 0.3$ between loci) across the six traits reached the genome-wide (GW) significance threshold of $P < 5 \times 10^{-8}$. The $-\log_{10}(P \text{ value})$ genome-wide association plots for the meta-analysis of each of the 6 traits are shown in Figure 1. Corresponding QQ-plots are shown in Supplementary Figure 1a and the genomic control lambda (λ_{GC}) values in Supplementary Table 2. The genomic control inflation factor post-meta-analysis, which was not corrected at the meta-analysis level, showed no systematic inflation (Hgb $\lambda_{GC} = 1.066$; Hct $\lambda_{GC} = 1.045$; MCH $\lambda_{GC} = 1.014$; MCHC $\lambda_{GC} = 0.995$; MCV $\lambda_{GC} = 1.029$; and RBC $\lambda_{GC} = 1.029$; Supplementary Table 2). The meta-analysis results for all traits are summarized in Table 2, which is organized by the 23 independent loci and includes gene annotation information for each locus. The table also lists for each trait the number of SNPs exceeding the GW significance level. Altogether, there were 45 trait-locus combinations with at least one GW significant SNP. The complete set of SNP associations identified by the CHARGE meta-analysis is provided in Supplementary Table 3. Replication and further analysis focused on the 45 SNPs that gave the smallest P values for each of the 45 trait-locus findings in CHARGE.

Independent replication

Replication of the 45 SNPs was conducted using a meta-analysis of association data in 9,456 independent European-ancestry individuals from five population-based cohorts in the HaemGen Consortium (Supplementary Note). A joint analysis of the HaemGen and CHARGE data showed a decrease in P values for all but two SNPs selected for replication. For one of the two SNPs (rs1800562) that did not show an improvement in P value when associated with Hct, the association to the Hgb trait was significant after Bonferroni correction, and for the second SNP (rs4466998), the association to MCV in the joint analysis of CHARGE and HaemGen data remained genome wide significant ($P = 4.91 \times 10^{-8}$). Significant independent replication for at least one trait was observed at 13 of 23 loci, using a Bonferroni-corrected significance threshold of $P < 0.0011$, or $0.05/45$. Taking the joint

meta-analysis results in sum, these data provide supportive evidence that the 23 loci from the discovery meta-analysis are true positives. Table 3 provides the full replication results, including beta coefficients, standard errors, and *P* values for the primary CHARGE findings, the HaemGen replication, and a combined meta-analysis of the two consortia for the 45 CHARGE trait-locus SNPs.

For each lead SNP in the 23 independent loci, percent variance explained for each of the lead SNPs in the corresponding trait is provided in Table 3, averaging the percent variance explained for each SNP across the CHARGE cohorts. The combination of lead SNPs from each of the trait loci showed that average percent variance explained by the combination of lead SNPs, beyond the variance explained by age and gender, was 1.14% of Hgb variation (rs10495928, rs1800562, rs10224002, rs16926246, rs11065987, rs6013509, rs855791); 1.16% of Hct variation (rs10168349, rs1800562, rs7385804, rs10224002, rs16926246, rs11065987, rs9483788, rs2413450); 4.53% of MCH variation (rs11915082, rs1408272, rs9349205, rs7776054, rs628751, rs10758658, rs1122794, rs11085824, rs2413450); 0.63% of MCHC variation (rs857721, rs9373124); 5.98% of MCV variation (rs2540917, rs9859260, rs172629, rs1800562, rs9349205, rs9374080, rs4895441, rs643381, rs12718597, rs7786877, rs10758658, rs11239550, rs4466998, rs7189020, rs7255045, rs2413450, rs131794); and 0.85% of RBC variation (rs9483788, rs2075671).

Annotation of associated loci

For 20 of the identified loci, top associated SNPs were identified within a +/- 60 Kb window of a RefSeq gene (Table 2). For three loci, chromosomes 4q12, 6q24.1, 20q13.2, no genes were identified within this window, with the nearest genes approximately 116 Kb, 89 Mb, and 50 Mb away, respectively. Of the 23 loci, previously reported mutations or genetic associations for erythrocyte traits, markers of iron status or fetal hemoglobin levels, have been noted at six loci containing the genes *HFE*, *TFR2*, *TMPRSS6*, *SPTA1*, *HBSIL-MYB*, and *BCL11A*. Most of the remaining loci have not previously been reported to be associated with erythrocyte traits, though several genes are known to have important roles in erythrocyte biology or erythropoiesis. Genes identified near the associated loci, and their associated erythrocyte traits, are presented in Table 2 and Figure 2. Gene annotations, including gene information, known genetic mutations causing hematologic and non-hematologic diseases, and previously defined roles in hematologic and cardiovascular systems are listed in Supplementary Table 4. We confirmed the association of the known C282Y (rs1800562) and H63D (rs1799945) mutations in the *HFE* gene, mutations that are already known to underlie hereditary hemochromatosis, with Hgb, Hct, MCH and MCV.

Gene expression in blood and endothelial cell lines

RNA expression levels for genes within a 1 Mb interval of each of the 23 loci we identified are presented in Supplementary Figure 2, for erythroid bodies (EBs), human umbilical vein endothelial cells (HUVECs), and seven other blood cell lines. For the top associated locus, chromosome 6q23.3, four genes were identified (*ALDH8A1*, *HBSIL*, *MYB*, *AH11*) within the 1 Mb interval. A heatmap showing gene expression levels for each of these four genes is shown in Figure 3, demonstrating approximately 2-fold expression of *MYB* in EBs compared to other cell lines. Gene expression was detected for genes in multiple loci

(Supplementary Figure 2). Since the identification of gene expression in biologically relevant tissues provides a rationale for prioritization of candidate genes for further genetic or functional investigations, we noted broad categories of expression patterns in EBs and HUVECs. The most highly expressed genes in EBs were in chromosomes 3q29 (*TFRC*), 6p22.2 (*HIST1H4C*, which is near *HFE*), 6p21.1 (*CCND3*), 10q21.3 (*HK1*), 12q24.12 (*RPL6P27*), 16p13.3 (*HBZ*, *HBA1*), and 22q12.3 (*TST*, *RAC2*). The most highly expressed genes in HUVECs were in chromosomes 6p22.2 (*HIST14HC*), 7q22.1 (*SERPINE1*), and 22q13.3 (*MFNG*).

Blood pressure analysis of SNPs associated with erythrocyte traits

The results of association testing for the 45 lead SNPs from the CHARGE analysis of erythrocyte traits within the 23 loci are summarized in Supplementary Table 5a. We identified associations at a Bonferroni-corrected significance threshold $P < 0.00135$ ($0.05/37$, since 37 of the 45 SNPs are unique) in chromosomes 12q24.1 (*SH2B3*) and 7q36.1 (*PRKAG2*). The previously reported association of the chromosome 12q24.1 (*SH2B3*) locus with systolic blood pressure (SBP) and diastolic blood pressure (DBP) was the most significant association (rs1106598, SBP $P = 1.2 \times 10^{-6}$, HTN $P = 0.0035$; rs1763023, DBP $P = 4.2 \times 10^{-8}$). In the reported BP and HTN analysis, signals at this locus spanned 700kb from rs3184504 to rs1106618815, and association signals in Hgb and Hct spanned 987kb, from rs3184504 in *SH2B3* to rs11066301 in *PTPN11* and contained multiple genes¹⁵. Inspection of RNA expression data (Supplementary Figure 2) showed that in this region, *SH2B3* and *ATXN2* show high levels of gene expression in erythroid bodies and endothelial cells. Nominal associations ($P < 0.05$) were identified in the chromosomes 6p22.2 (*HFE*), 6q24.1, 7q22.1 (*TFR2*), and 20q12.3 (Supplementary Table 5a), and results of the evaluation of SNPs associated with BP and hypertension are presented in Supplementary Table 5b.

Discussion

In this meta-analysis of genome-wide association data from 24,167 European-ancestry individuals from six cohort studies in the CHARGE Consortium, we identified 23 loci associated with at least one of the six erythrocyte traits Hgb, Hct, MCH, MCHC, MCV, and RBC. We sought evidence for replication in an independent analysis of data from 9,456 European-ancestry individuals in the HaemGen Consortium. In the joint meta-analysis, merging CHARGE and HaemGen data, all 23 loci had P values less than $P < 5 \times 10^{-8}$, implying strong associations that merit further study. Among the 23 loci, six were previously known QTLs, and 17 are novel loci, some of which contain genes known to be involved with iron homeostasis, erythropoiesis, globin synthesis and erythrocyte membrane function. Finally, an investigation of possible links between these erythrocyte traits and blood pressure and hypertension confirmed overlap at the previously known *SH2B3* locus, and identified additional suggestive associations, none of which met a genome-wide significance threshold.

The six erythrocyte traits studied included some that are highly correlated, and as expected, we observed a high degree of concordance in the results across the six traits. Among the many genome-wide significant associations identified, the patterns of association are likely

to reflect the correlations among these related traits. Interestingly, MCHC, a ratio of Hgb and Hct, two directly measured traits, is uniquely associated with chromosome 1q23.1 (*SPTA1*), a gene with several rare mutations known to cause deformation of erythrocytes^{16, 17}. Reviewing the results in total, we observed there were generally three patterns of significant associations among the six traits (Figure 2). Results were generally similar for: (1) Hgb and Hct, which are mainly quantitative measures of hemoglobin in the blood; (2) MCH and MCV, representing erythrocyte size and quantity of hemoglobin per erythrocyte; and (3) MCHC, the ratio of Hgb to Hct, which appears somewhat distinct from the other traits. Across the six traits studied, the strongest signal was found in the *HBSIL/MYB* locus on chromosome 6q23, which was observed for five of the six individual traits (Hct, MCH, MCHC, MCV, RBC) at the genome-wide significance level. This locus also provided a modest but non-significant result for Hgb (rs4895441, $P = 4.8 \times 10^{-4}$). In addition, the Hgb/Hct and MCH/MCV patterns overlapped for associations in the chromosome 6p22.2 (*HFE*), 22q12.3 (*TMPRSS6*) and 7q22.1 (*TFR2/EPO*) loci. The RBC results represent a subset of the overlap between the Hgb/Hct and MCH/MCV patterns, with associations observed in the 7q22.1 (*TFR2/EPO*) and 6q23 (*HBSIL/MYB*) loci. Across the erythrocyte traits, overlap occurs where known patterns of traits characterize various clinically observed hematologic diseases, providing a possible context in which to interpret the overlap of associations.

We annotated and categorized the findings of our analyses by association with known genetic disorders, biologic function, or altered function of the hematopoietic system, to assist with interpretation of the findings (Supplementary Table 4). We here consider these multiple findings in light of their potential role in several processes critical to erythrocyte biology, including iron homeostasis, erythrocyte membrane function, erythropoiesis and globin synthesis.

Iron homeostasis

We identified genome-wide significant association of SNPs within the *HFE* gene with Hgb, Hct, MCH, and MCV. C282Y mutation in *HFE* is the principal cause of hereditary hemochromatosis, a common autosomal recessive iron overload disease in individuals of northern European descent¹⁸. This mutation was associated with increased MCV and Hgb concentrations in a study of individuals drawn from a hemochromatosis and iron overload screening study⁷, and this variant was the lead association result for both Hgb and Hct in our study. Heterozygotes for either allele do not manifest clinical iron overload but may display an increased iron uptake and “resistance” to anemia, and the C282Y mutation may increase risk of coronary heart disease by increasing iron stores and lipid oxidation^{19, 20}. The *HFE* gene induces expression of the iron regulatory hormone hepcidin. Hepcidin has recently emerged as the likely link between the inflammatory response and the handling of iron for erythropoiesis by both downregulating the absorption of iron in the intestine and by inhibiting the release of iron from macrophages²¹⁻²⁴. SNPs within the *TMPRSS6* gene were associated with Hgb, Hct, MCH and MCV. *TMPRSS6* was identified by linkage and association studies in five families and two sporadic cases with iron-refractory iron deficiency anemia, a rare Mendelian disease²⁵. *TMPRSS6* encodes a type II transmembrane serine protease produced by the liver that regulates the expression of hepcidin²⁵. The

transferrin receptor (encoded by *TFRC(TFRI)*) and transferrin receptor 2 (*TFR2*) are highly homologous type II trans-membrane proteins in the transferrin protein family. SNPs within *TFRC* were associated with MCH and MCV, and SNPs within *TFR2* were associated with Hct and MCV. Reduced *TFRC* expression is associated with anemia²⁶. Existing evidence indicates that *TFR2* is also a modulator of hepcidin expression, and mutations in *TFR2* cause hemochromatosis type 327.

Erythropoiesis and globin synthesis

Two loci, chromosome 2p16.1 (*BCL11A*) associated with MCV, and chromosome 6q23.3 (*HBSIL-MYB*) associated with all traits except for Hgb, are related to variation in fetal hemoglobin and hemoglobin beta levels^{28, 29}. *BCL11A* is an oncogene related to B-cell malignancies³⁰ and regulates fetal hemoglobin expression³¹. *BCL11A* is expressed in erythroid precursors, and we observed *BCL11A* expression in EBs (Supplementary Figure 2) making it a biologically plausible candidate gene for erythrocyte trait variation³². A healthy population study showed polymorphisms in *HBSIL* and *MYB* influences erythrocyte, platelet, and monocyte counts¹⁰. Although the role of *HBSIL* is unknown, *MYB* has been associated with proliferation, survival, and differentiation of hematopoietic progenitor cells^{33, 34}. *MYB* is also associated with eosinophil counts in blood and atopic asthma³⁵. There are strong associations between SNPs within this locus and multiple erythrocyte traits (lead SNP rs4895441, Hgb $P = 4.8 \times 10^{-4}$; Hct $P = 9.7 \times 10^{-10}$; MCH $P = 7.8 \times 10^{-32}$; MCHC $P = 4.5 \times 10^{-9}$; MCV $P = 1.0 \times 10^{-57}$; and RBC $P = 2.2 \times 10^{-15}$). These strong genetic effects may explain why several prior linkage analyses of erythrocyte traits have identified this chromosomal region^{6, 11, 13}. SNPs within *SH2B3* are associated with Hct and Hgb. Interestingly, the SNPs within this gene are associated with blood pressure, myocardial infarction, type 1 diabetes, and celiac disease^{15, 35-39}. *SH2B3* is expressed in hematopoietic precursor cells and increases hematopoietic progenitors of erythroid, megakaryocytic, and myeloid lineages^{35, 40}. *SH2B3* is also expressed in vascular endothelium, where it promotes inflammation and may thereby contribute to vascular disease. Expression in different cell lineages and tissues may underlie the diverse pleiotropic consequences of *SH2B3* on hematopoietic traits, autoimmune diseases, and vascular diseases. In the same locus, the *PTPN11* gene product interacts with the transcription factor SHP2, which has an essential role in blood development that has been demonstrated in a murine *Shp2*^{-/-} model⁴¹, and *PTPN11* mutations cause Noonan's and LEOPARD syndromes and juvenile myelomonocytic leukemia⁴²⁻⁴⁴. Lastly, we identified associations for Hct, MCV and RBC near the *EPO* gene. Erythropoietin, a glycoprotein hormone that controls erythropoiesis, is the first human recombinant hematopoietic protein approved for human use and is now used widely for the treatment of anemia. *EPO* variants have previously been described in association with diabetic retinal and renal vascular complications⁴⁵ but not with erythrocyte traits.

Erythrocyte membrane

SNPs within the *SPTA1* gene were associated with MCHC. *SPTA1* encodes erythroid spectrin, a protein in the erythrocyte membrane, and is essential in determining the shape and deformability of erythrocytes. Spectrin mutations have been previously associated with

hemolytic anemia, elliptocytosis, spherocytosis, and propoikilocytosis, but not with variations in MCHC outside of disease states^{16, 17}.

Gene expression

The gene expression data may be viewed as additional annotation of the 23 loci we identified, confirming which genes are expressed in cell types of interest. These data may be used to generate further specific hypotheses that can then be tested at a functional and molecular level.

Multiple lines of evidence formed the basis for our rationale to study the relationship of the SNPs identified through the study of erythrocyte traits to blood pressure (BP) and hypertension. Prior studies have shown that Hgb and Hct levels are associated with increased risk for hypertension and a variety of other vascular diseases and mortality^{1-3, 46-49}. From a rheologic perspective, blood viscosity depends largely on Hgb or Hct levels and is a determinant of blood pressure⁵⁰⁻⁵³. There is an inverse relationship between viscosity and vascular blood flow⁵⁴, and elevated Hct thereby hampers organ perfusion. Given these findings, we are intrigued by the overlap between the association results for Hgb and Hct from our study and the recently reported associations for BP and hypertension^{15, 36}. We observed overlap of associations in the chromosome 12q24.12 region across a 987 kb linkage disequilibrium block, containing SH2B3, ATXN2, BRAP, C12orf03, TRAFD1, ACAD10, TMEM116 and *PTPN11*. We also identified associations within the 7q36.1 region containing *PRKAG2*, which does not have specifically known hematologic or vascular roles, but mutations in this gene cause cardiomyopathy and cardiac conduction system disorders^{55, 56}. Neither causality nor independence of these associations is necessarily supported by these findings. However, these associations suggest that common genetic bases may underlie some of the correlation seen between erythrocyte, BP and hypertension traits. Further confirmation in large independent cohorts may provide stronger evidence for the strength and consistency of the associations with hypertension.

Limitations of this study include restriction of the discovery and replication analyses to individual of European ancestry. Several spectrin and globin mutations have been identified in African American kindreds and the prevalence of hemoglobinopathies of various types is generally higher among individuals of African ancestry, highlighting the need for further investigation of these findings in individuals of non-European-ancestry⁵⁷. As with any meta-analysis of genome-wide association results across different cohorts, population structure and other sources of heterogeneity may have caused false positives or false negatives. To assess population structure, we examined the per-cohort λ_{GC} , demonstrating that these values were consistently below 1.08, and we applied genomic control to the cohort-level test statistics. The final meta-analyses also showed no systematic inflation of the distribution of the final association statistics. In the replication analysis, for those loci that did not meet a conservative replication test, power may have been limited, and many are likely to improve with additional study. Finally, the interpretation of multiple analyses of correlated traits requires caution, particularly in attributing causality or independence of effects. We take the findings from our analyses of the six erythrocyte traits to indicate a set of loci that are of interest with regards to erythrocyte production, homeostasis and function.

Specific differences in association patterns may highlight different pathways, and to understand this more deeply, further studies are needed.

In summary, we have identified and validated common variants at several known and novel loci that influence the levels of six clinically relevant red blood cell measures in population-based cohorts. These QTLs have implications for understanding a variety of hematologic diseases as well as correlates of erythrocyte traits, such as BP and hypertension. Further studies are warranted to define these variants in the extremes of the distributions of these traits and ethnically diverse populations and to understand the functional impact of variants at the implicated candidate genes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Santhi K. Ganesh^{*,1,2,3}, Neil A. Zakai^{*,4}, Frank J A. van Rooij^{*,5,6}, Nicole Soranzo^{*,7,8}, Albert V. Smith^{*,9}, Michael A. Nalls^{*,10}, Ming-Huei Chen^{11,12}, Anna Kottgen¹³, Nicole L. Glazer^{14,15}, Abbas Dehghan^{5,6}, Brigitte Kuhnel¹⁶, Thor Aspelund^{9,17}, Qiong Yang^{11,18}, Toshiko Tanaka^{19,20}, Andrew Jaffe¹³, Joshua C M. Bis¹⁵, Germaine C. Verwoert^{5,6,21}, Alexander Teumer²², Caroline S. Fox^{2,11}, Jack M. Guralnik²³, Georg B. Ehret^{3,24}, Kenneth Rice²⁵, Janine F. Felix^{5,6}, Augusto Rendon²⁶, Gudny Eiriksdottir⁹, Daniel Levy^{2,11}, Kushang V. Patel²³, Eric Boerwinkle²⁷, Jerome I. Rotter²⁸, Albert Hofman^{5,6}, Jennifer G. Sambrook²⁶, Dena Hernandez²⁹, Gang Zheng³⁰, Stefania Bandinelli³¹, Andrew B. Singleton¹⁰, Josef Coresh¹³, Thomas Lumley²⁵, André G. Uitterlinden^{5,6,21,32}, Janine M. vanGils³³, Lenore J. Launer²³, L. Adrienne Cupples^{11,18}, Ben A. Oostra^{5,6,34}, Jaap-Jan Zwaginga^{35,36}, Willem H. Ouwehand^{26,7}, Swee-Lay Thein³⁷, Christa Meisinger¹⁶, Panos Deloukas⁷, Matthias Nauck³⁸, Tim Spector⁸, Christian Gieger¹⁶, Vilmundur Gudnason^{9,17}, Cornelia M. van Duijn^{5,6}, Bruce M. Psaty^{14,15,39,40,41}, Luigi Ferrucci¹⁹, Aravinda Chakravarti³, Andreas Greinacher^{^,42}, Christopher J. O'Donnell^{^,2,11,43}, Jacqueline C M. Witteman^{^,5,6}, Susan Furth^{^,44}, Mary Cushman^{^,4}, Tamara B. Harris^{^,23}, and Jing-Ping Lin^{^,30}

Affiliations

¹National Human Genome Research Institute, Division of Intramural Research, Bethesda, MD, USA ²National Heart, Lung and Blood Institute, Division of Intramural Research, Bethesda, MD, USA ³McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA ⁴Departments of Medicine and Pathology, University of Vermont, Burlington, VT, USA ⁵Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands ⁶The Netherlands Genomics Initiative-sponsored Netherlands Consortium for Healthy Aging (NGINCHA) ⁷Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, CB10 1SA, UK ⁸Department of Twin Research and Genetic Epidemiology, King's College London, Lambeth Palace Rd, London, SE1 7EH, UK ⁹Icelandic Heart

Association, Kopavogur, Iceland ¹⁰Laboratory of Neurogenetics, National Institute of Aging, NIH, Bethesda, MD, USA ¹¹NHLBI's Framingham Heart Study, Framingham, MA, USA ¹²Department of Neurology, Boston University School of Medicine, Boston, MA, USA ¹³Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA ¹⁴Department of Medicine, University of Washington, Seattle, WA, USA ¹⁵Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA ¹⁶Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, 85764, Germany ¹⁷University of Iceland, Reykjavik, Iceland ¹⁸Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA ¹⁹Longitudinal Studies Section - Clinical Research Branch., National Institute of Aging, NIH, Baltimore, MD, USA ²⁰MedStar Research Institute, Hyattsville, MD, USA ²¹Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands ²²Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-Universität Greifswald, 17487 Greifswald, Germany ²³Laboratory for Epidemiology, Demography, and Biometry, National Institute of Aging, NIH, Bethesda, MD, USA ²⁴Cardiology Center, Geneva University Hospital, Switzerland ²⁵Department of Biostatistics, University of Washington, Seattle, WA, USA ²⁶Department of Haematology, University of Cambridge and National Health Service Blood and Transplant, Long Road, Cambridge, CB2 0PT, UK ²⁷Human Genetics Center, University of Texas Health Science Center, Houston, TX, USA ²⁸Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA ²⁹Intramural Research Program, Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD, USA ³⁰Division of Cardiovascular Sciences, Office of Biostatistical Research, National Heart, Lung and Blood Institute, National Institutes of Health, USA ³¹Geriatric Unit ASF, Firenze, Italy ³²Department of Clinical Chemistry, Erasmus MC, Rotterdam, the Netherlands ³³Leiden University Medical Center, Leiden, The Netherlands ³⁴Department of Clinical Genetics, Erasmus MC, Rotterdam, the Netherlands ³⁵Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, Netherlands ³⁶Department of Experimental Immunohematology, Sanquin Research, Amsterdam, The Netherlands ³⁷Department of Haematology, King's College London, London SE5 9NU ³⁸Institut für Klinische Chemie und Laboratoriumsmedizin, Ernst-Moritz-Arndt Universität Greifswald, Greifswald, 17487, Germany ³⁹Department of Epidemiology, University of Washington, Seattle, WA, USA ⁴⁰Department of Health Services, University of Washington, Seattle, WA, USA ⁴¹Group Health Research Institute, Group Health, Seattle, WA, USA ⁴²Institut für Immunologie und Transfusionsmedizin, Ernst-Moritz-Arndt-Universität Greifswald, Greifswald, D-17487, Germany ⁴³Cardiology Division, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA ⁴⁴Department of Pediatrics, Johns Hopkins School of Medicine, Baltimore, MD, USA

Acknowledgements

Complete study acknowledgments are listed in the Supplementary Note. The authors thank the studies' participants, staff and the funding agencies for their support.

Appendix

Methods

CHARGE Consortium organization and study samples

We performed a cross-sectional analysis of genotype and phenotypic data on erythrocyte traits in the CHARGE consortium¹⁴, which includes five cohort studies that have genotyped high density SNP markers and have phenotypic data on erythrocyte traits (AGES N=3,205; ARIC N=7,803; CHS N=3,256; FHS N=3,359; RS N=5,523) and InCHIANTI (N=1,021). Each participating study was approved by its corresponding Institutional Review Board, and all study subjects provided informed consent for participation in the study and genetic research. Participants were excluded if they were of non-European ancestry, as determined by self-report, and also by principal component analysis in ARIC and RS. Detailed methods for each of the participating cohorts are provided in the Supplementary Note.

Genotyping and imputation

Briefly, each study genotyped samples using high-density SNP marker platforms (Affymetrix SNP6.0 - ARIC; Affymetrix 500K - FHS; Illumina 370K - AGES, CHS; Illumina 550K-InCHIANTI, RS). Genotypes were then imputed to a set of approximately 2.5M HapMap SNPs using Phase II CEU HapMap individuals for reference using either MACH (<http://www.sph.umich.edu/csg/abecasis/MACH>) (ARIC, AGES, FHS, InCHIANTI, RS) or BimBam58 (CHS) software.

Erythrocyte phenotypes

Erythrocyte parameters studied were 1) hemoglobin concentration (Hgb), the concentration of hemoglobin within whole blood; 2) hematocrit (Hct), the percent of whole blood comprised of cellular erythrocyte elements 3) red blood cell count (RBC), the number of red blood cells per volume of blood. 4) mean corpuscular volume (MCV), the average erythrocyte volume; 5) mean corpuscular hemoglobin (MCH), the average quantity of Hgb per erythrocyte; and 6) mean corpuscular hemoglobin concentration (MCHC), the ratio of Hgb to Hct. The definition and units of each trait are provided in Supplementary Table 1. Blood was drawn from each participant using standard phlebotomy methods, and erythrocyte measures were obtained using standard clinical assays in certified laboratories.

Phenotype modeling and statistical analysis

All traits studied were continuous. Based on prior convention and visual inspection of the data, MCH, MCHC and MCV were natural log transformed, RBC was square root transformed, and Hgb and Hct were not transformed prior to analyses. In order to focus on determinants of variation of these traits in the general population, rather than on specific hematologic diseases which are over-represented at the tails of the distribution for each of the traits, we restricted analysis to those individuals within three standard deviations of the sample mean within each cohort. For each SNP meeting QC criteria, linear regression was used to assess association with trait, separately for all six traits. An additive genetic model was used throughout. These regressions were adjusted for age, gender, and site in the multi-center cohorts. In FHS, linear mixed effects models were used to account for relatedness,

and these models included adjustment for principal components, computed using Eigenstrat 2.059. The genome-wide level of significance threshold was set at $P < 5 \times 10^{-8}$.

Each cohort's results were forwarded to a central repository, including regression coefficients, standard errors, sample size, imputation quality, minor allele designation and minor allele frequency. Prior to meta-analysis, we performed genomic control on each cohort-specific distribution of the association test statistics for each trait⁶⁰. We also filtered out SNPs with allelic frequency less than 1% or poor imputation quality (the ratio of observed variance of imputed allele counts to the expected variance of imputed allele counts > 1.1 from the imputation software output). Separately for each SNP and trait, within-cohort association results were combined in an inverse variance weighted meta-analysis, as implemented in METAL (<http://www.sph.umich.edu/csg/abecasis/Metal/index.html>). After meta-analysis, genomic control inflation factor (λ_{GC}) was again calculated to assess stratification between the cohorts and resulting inflation of the test statistics. Genomic control was not applied to the final meta-analysis results. The SNAP program (<http://www.broadinstitute.org/mpg/snap>) was used to estimate linkage disequilibrium between the associated loci. Percent variance explained within each cohort was calculated from the r^2 estimate derived from a linear regression model for individual lead SNPs at each trait locus and the combination of all leads SNPs per trait, accounting for age and gender as well.

Replication samples and analysis

The replication set included samples from the five population-based cohorts of individuals of European ancestry that comprise the HaemGen Consortium (Study of Health in Pomerania (SHIP) N=3,200; UK Blood Services Common Control (UKBS1-CC1) N=1,290; Twins UK adult twin registry N=1,510; KORAF3 500K study population N =1,643; and KORA-F4 N=1,814). Further information on these cohorts is provided in the Supplementary Note. Trait definitions were identical to our initial study. Analysis for each of the six traits undergoing replication was implemented in the same way as for the CHARGE analyses. For replication, we used a Bonferroni correction for the number of SNPs tested by the HaemGen Consortium. Given the smaller sample size available for the HaemGen replication analysis relative to the CHARGE discovery analysis, we performed a combined meta-analysis on the top trait-locus SNPs identified by the CHARGE meta-analysis, to assess impact on the association signals.

Gene expression in blood cell subtypes and endothelial cells

To assist with prioritization of the candidate genes identified in this study, we evaluated RNA expression in eight blood cell lines (stem-cell derived erythroblasts (EB) and megakaryocytes (MK), CD14+ monocytes, CD56+ natural killer (NK) cells, CD4+ T helper (Th) lymphocytes, CD8+ cytotoxic T (Tc) lymphocytes, and CD66b+ granulocytes), using an established catalogue of gene expression in these lineages⁶¹; we also examined gene expression in cultured human umbilical vein endothelial cells (HUVECs)⁶². For each of the 23 identified loci, we examined the transcript levels in the nine cell types of all genes within a ± 500 kb window centered on the lead association SNP.

Briefly, EB and MK samples were obtained by culturing of CD34-positive haematopoietic stem cells purified from cord blood cultured with EPO-IL3-SCF for 10 days and with TPO, respectively⁶¹. Cultured cells were sorted by fluorescence-activated cell sorting using either a monoclonal antibody against CD235a (glycophorin A) or against CD41 (α IIb integrin). The other six blood cell types were purified from the peripheral blood of seven healthy subjects using the corresponding CD markers. Blood cell types were hybridized onto Illumina V2 Ref6 gene expression beadarrays, and the detailed methods for isolation of other hematologic cell types, RNA extraction and microarray analysis are described elsewhere⁶¹. Cultured human umbilical vein endothelial cells (HUVECs) and expression profiles were ascertained as has been described elsewhere⁶² and using the same microarray platform⁶¹.

Raw data was VST transformed using the R/Bioconductor package “lumi,”⁶³ followed by quantile normalization. The mode of the transformed signal intensity is 7.94 and can be taken as the background intensity. The maximum probe intensity was 15.85 corresponding to a signal intensity of 59,000 in a linear scale. For each cell type and probe, a Grubbs test for outlier identification was used and samples with *P* values less than 0.01 were removed (implemented in the R package “outliers”). This was not performed for the MK, EB or HUVEC cell types due to their smaller sample size (4, 4, and 3 respectively). Probe mappings were obtained from re-annotation efforts available at <http://www.compbio.group.cam.ac.uk/Resources/Annotation/>.

Annotation of genetic loci

To determine genes within or neighboring each locus, we examined RefSeq gene annotations, build 36. To annotate the loci and genes, we reviewed the literature, OMIM⁶⁴ and the Genetic Association Database⁶⁵ (Supplementary Table 4).

Blood pressure and hypertension analyses

Following the observation that epidemiologic studies have identified a strong yet unexplained link between Hgb and Hct and blood pressure, we examined the associations of these SNPs to those previously reported meta-analysis within the CHARGE consortium for systolic blood pressure (SBP), diastolic blood pressure (DBP) and hypertension^{1, 2, 15}. The CHARGE cohorts that contributed samples to the BP and hypertension analyses were AGES, ARIC, CHS, FHS, and RS, with a total sample size of $N=29,136$. SBP and DBP measured at the first visit attended were continuous traits, and hypertension was analyzed as a dichotomous trait¹⁵. We tested the association of the lead SNP per locus for each trait (45 SNPs) for association with SBP, DBP and hypertension¹⁵, using a Bonferroni corrected threshold for the number of SNPs tested. We additionally reversed the analysis, examining the test statistics within each of the six erythrocyte traits for those SNPs reported to be associated with SBP, DBP or hypertension¹⁵.

References

1. Havlik RJ, et al. Evidence for additional blood pressure correlates in adults 20-56 years old. *Circulation*. 1980; 61:710–5. [PubMed: 7357712]

2. Tell GS, et al. The Cardiovascular Health Study; Cardiovascular Health Study (CHS) Collaborative Research Group. Correlates of blood pressure in community-dwelling older adults. *Hypertension*. 1994; 23:59–67. [PubMed: 8282331]
3. Zakai NA, et al. A prospective study of anemia status, hemoglobin concentration, and mortality in an elderly cohort: the Cardiovascular Health Study. *Arch Intern Med*. 2005; 165:2214–20. [PubMed: 16246985]
4. Whitfield JB, Martin NG. Genetic and environmental influences on the size and number of cells in the blood. *Genet Epidemiol*. 1985; 2:133–44. [PubMed: 4054596]
5. Evans DM, Frazer IH, Martin NG. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res*. 1999; 2:250–7. [PubMed: 10723803]
6. Lin JP, et al. Evidence for linkage of red blood cell size and count: genome-wide scans in the Framingham Heart Study. *Am J Hematol*. 2007; 82:605–10. [PubMed: 17211848]
7. McLaren CE, et al. Determinants and characteristics of mean corpuscular volume and hemoglobin concentration in white HFE C282Y homozygotes in the hemochromatosis and iron overload screening study. *Am J Hematol*. 2007; 82:898–905. [PubMed: 17597476]
8. Zeng SM, Yankowitz J, Widness JA, Strauss RG. Etiology of differences in hematocrit between males and females: sequence-based polymorphisms in erythropoietin and its receptor. *J Genet Specif Med*. 2001; 4:35–40. [PubMed: 11324238]
9. Zeng SM, Yankowitz J, Widness JA, Strauss RG. Sequence-based polymorphisms in members of the apoptosis Bcl-2 gene family and their association with hematocrit level. *J Genet Specif Med*. 2003; 6:36–42. [PubMed: 14714449]
10. Menzel S, et al. The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. *Blood*. 2007; 110:3624–6. [PubMed: 17712044]
11. Iliadou A, et al. Genomewide scans of red cell indices suggest linkage on chromosome 6q23. *J Med Genet*. 2007; 44:24–30. [PubMed: 16950815]
12. Yang Q, Kathiresan S, Lin JP, Tofler GH, O'Donnell CJ. Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. *BMC Med Genet*. 2007; 8(Suppl 1):S12. [PubMed: 17903294]
13. Lin JP, O'Donnell CJ, Levy D, Cupples LA. Evidence for a gene influencing haematocrit on chromosome 6q23-24: genomewide scan in the Framingham Heart Study. *J Med Genet*. 2005; 42:75–9. [PubMed: 15635079]
14. Psaty BM, O'Donnell CJ, Gudnason V. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circ Cardiovascular Genetics*. 2008 al., e. in press.
15. Levy D, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009
16. Agre P, Casella JF, Zinkham WH, McMillan C, Bennett V. Partial deficiency of erythrocyte spectrin in hereditary spherocytosis. *Nature*. 1985; 314:380–3. [PubMed: 3982506]
17. Marchesi SL, et al. Mutant forms of spectrin alpha-subunits in hereditary elliptocytosis. *J Clin Invest*. 1987; 80:191–8. [PubMed: 3597773]
18. Feder JN, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet*. 1996; 13:399–408. [PubMed: 8696333]
19. Roest M, et al. Heterozygosity for a hereditary hemochromatosis gene is associated with cardiovascular death in women. *Circulation*. 1999; 100:1268–73. [PubMed: 10491369]
20. Tuomainen TP, et al. Increased risk of acute myocardial infarction in carriers of the hemochromatosis gene Cys282Tyr mutation: a prospective cohort study in men in eastern Finland. *Circulation*. 1999; 100:1274–9. [PubMed: 10491370]
21. Vecchi C, et al. ER stress controls iron metabolism through induction of hepcidin. *Science*. 2009; 325:877–80. [PubMed: 19679815]
22. Swinkels DW, Wetzels JF. Hepcidin: a new tool in the management of anaemia in patients with chronic kidney disease? *Nephrol Dial Transplant*. 2008; 23:2450–3. [PubMed: 18495744]
23. Roy CN, et al. Hepcidin antimicrobial peptide transgenic mice exhibit features of the anemia of inflammation. *Blood*. 2007; 109:4038–44. [PubMed: 17218383]

24. Feder JN, et al. The hemochromatosis gene product complexes with the transferrin receptor and lowers its affinity for ligand binding. *Proc Natl Acad Sci U S A*. 1998; 95:1472–7. [PubMed: 9465039]
25. Finberg KE, et al. Mutations in *TMPRSS6* cause iron-refractory iron deficiency anemia (IRIDA). *Nat Genet*. 2008; 40:569–71. [PubMed: 18408718]
26. Kerényi MA, et al. *Stat5* regulates cellular iron uptake of erythroid cells via *IRP-2* and *TfR-1*. *Blood*. 2008; 112:3878–88. [PubMed: 18694996]
27. Roetto A, et al. Hemochromatosis due to mutations in transferrin receptor 2. *Blood Cells Mol Dis*. 2002; 29:465–70. [PubMed: 12547237]
28. Uda M, 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; 105:1620–5. [PubMed: 18245381]
29. Menzel S, et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet*. 2007; 39:1197–9. [PubMed: 17767159]
30. Bea S, et al. Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood*. 2005; 106:3183–90. [PubMed: 16046532]
31. Sankaran VG, et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor *BCL11A*. *Science*. 2008; 322:1839–42. [PubMed: 19056937]
32. Higgs DR, Wood WG. Genetic complexity in sickle cell disease. *Proc Natl Acad Sci U S A*. 2008; 105:11595–6. [PubMed: 18695233]
33. Sandberg ML, et al. *c-Myb* and *p300* regulate hematopoietic stem cell proliferation and differentiation. *Dev Cell*. 2005; 8:153–66. [PubMed: 15691758]
34. Emambokus N, et al. Progression through key stages of haemopoiesis is dependent on distinct threshold levels of *c-Myb*. *Embo J*. 2003; 22:4478–88. [PubMed: 12941699]
35. Gudbjartsson DF, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet*. 2009; 41:342–7. [PubMed: 19198610]
36. Newton-Cheh C, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet*. 2009
37. Todd JA, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet*. 2007; 39:857–64. [PubMed: 17554260]
38. Hunt KA, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet*. 2008; 40:395–402. [PubMed: 18311140]
39. Smyth DJ, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med*. 2008; 359:2767–77. [PubMed: 19073967]
40. Velazquez L, et al. Cytokine signaling and hematopoietic homeostasis are disrupted in *Lnk*-deficient mice. *J Exp Med*. 2002; 195:1599–611. [PubMed: 12070287]
41. Saxton TM, et al. Abnormal mesoderm patterning in mouse embryos mutant for the SH2 tyrosine phosphatase *Shp-2*. *Embo J*. 1997; 16:2352–64. [PubMed: 9171349]
42. Jongmans M, et al. Genotypic and phenotypic characterization of Noonan syndrome: new data and review of the literature. *Am J Med Genet A*. 2005; 134A:165–70. [PubMed: 15723289]
43. Kalidas K, et al. Genetic heterogeneity in LEOPARD syndrome: two families with no mutations in *PTPN11*. *J Hum Genet*. 2005; 50:21–5. [PubMed: 15690106]
44. Tartaglia M, et al. Somatic mutations in *PTPN11* in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat Genet*. 2003; 34:148–50. [PubMed: 12717436]
45. Tong Z, et al. Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications. *Proc Natl Acad Sci U S A*. 2008; 105:6998–7003. [PubMed: 18458324]
46. Wannamethee G, Shaper AG. Haematocrit: relationships with blood lipids, blood pressure and other cardiovascular risk factors. *Thromb Haemost*. 1994; 72:58–64. [PubMed: 7974376]
47. Gagnon DR, Zhang TJ, Brand FN, Kannel WB. Hematocrit and the risk of cardiovascular disease--the Framingham study: a 34-year follow-up. *Am Heart J*. 1994; 127:674–82. [PubMed: 8122618]

48. Elwood PC, Waters WE, Benjamin IT, Sweetnam PM. Mortality and anaemia in women. *Lancet*. 1974; 1:891–4. [PubMed: 4133418]
49. Sarnak MJ, et al. Anemia as a risk factor for cardiovascular disease in The Atherosclerosis Risk in Communities (ARIC) study. *J Am Coll Cardiol*. 2002; 40:27–33. [PubMed: 12103252]
50. de Simone G, et al. Association of blood pressure with blood viscosity in american indians: the Strong Heart Study. *Hypertension*. 2005; 45:625–30. [PubMed: 15699438]
51. Letcher RL, Chien S, Pickering TG, Laragh JH. Elevated blood viscosity in patients with borderline essential hypertension. *Hypertension*. 1983; 5:757–62. [PubMed: 6352482]
52. Sharp DS, et al. Mean red cell volume as a correlate of blood pressure. *Circulation*. 1996; 93:1677–84. [PubMed: 8653873]
53. Wells RE Jr, Merrill EW. Influence of flow properties of blood upon viscosity-hematocrit relationships. *J Clin Invest*. 1962; 41:1591–8. [PubMed: 14040228]
54. Ott EO, Lechner H, Aranibar A. High blood viscosity syndrome in cerebral infarction. *Stroke*. 1974; 5:330–3. [PubMed: 4836535]
55. Gollob MH, et al. Identification of a gene responsible for familial Wolff-Parkinson-White syndrome. *N Engl J Med*. 2001; 344:1823–31. [PubMed: 11407343]
56. Gollob MH, et al. Novel PRKAG2 mutation responsible for the genetic syndrome of ventricular preexcitation and conduction system disease with childhood onset and absence of cardiac hypertrophy. *Circulation*. 2001; 104:3030–3. [PubMed: 11748095]
57. Beutler E, Waalen J. The definition of anemia: what is the lower limit of normal of the blood hemoglobin concentration? *Blood*. 2006; 107:1747–50. [PubMed: 16189263]
58. Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet*. 2007; 3:e114. [PubMed: 17676998]
59. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–9. [PubMed: 16862161]
60. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999; 55:997–1004. [PubMed: 11315092]
61. Watkins NA, et al. A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood*. 2009; 113:e1–9. [PubMed: 19228925]
62. da Costa Martins P, et al. P-selectin glycoprotein ligand-1 is expressed on endothelial cells and mediates monocyte adhesion to activated endothelium. *Arterioscler Thromb Vasc Biol*. 2007; 27:1023–9. [PubMed: 17322099]
63. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008; 24:1547–8. [PubMed: 18467348]
64. Online Mendelian Inheritance in Man, OMIM (TM). McKusick-Nathans Institute of Genetic Medicine. Johns Hopkins University; National Center for Biotechnology Information, National Library of Medicine; Baltimore, MD: Bethesda, MD: date of download
65. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. *Nat Genet*. 2004; 36:431–2. [PubMed: 15118671]

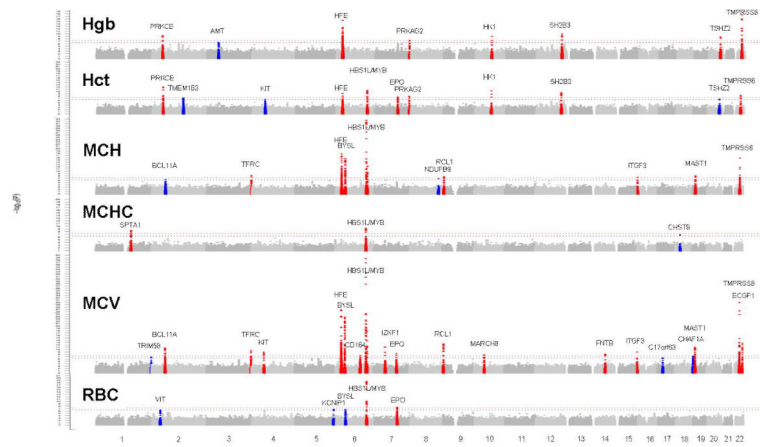


Figure 1. Overview of CHARGE meta-analysis results for six erythrocyte traits: hemoglobin concentration (Hgb), hematocrit (Hct), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), erythrocyte count (RBC). $-\log_{10}(P)$ value is plotted on the y-axis against genomic position of each SNP. Genomic loci with significant association ($P < 5 \times 10^{-8}$) are plotted in red, and loci with suggestive evidence are in blue ($P < 4 \times 10^{-7}$).

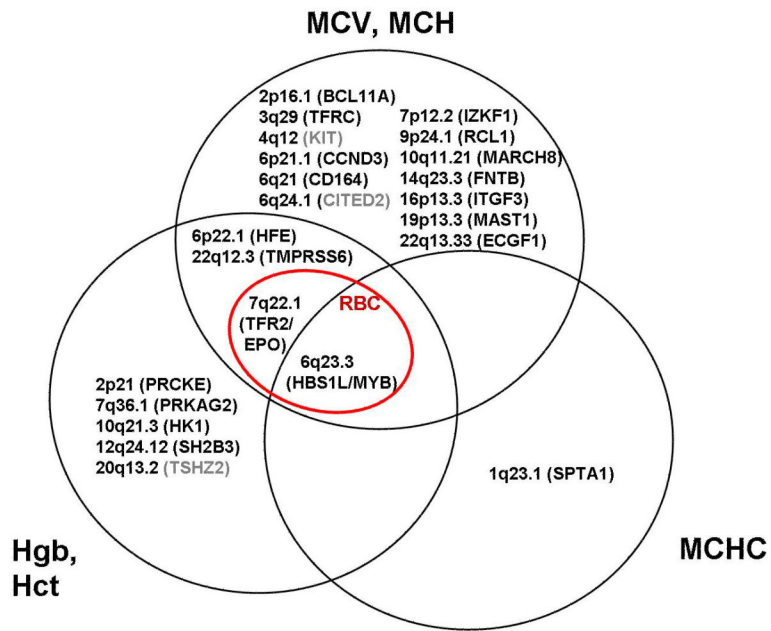


Figure 2. Results of the CHARGE meta-analysis are organized into a Venn diagram, demonstrating overlap of loci meeting a genome-wide significance threshold of $P < 5 \times 10^{-8}$.

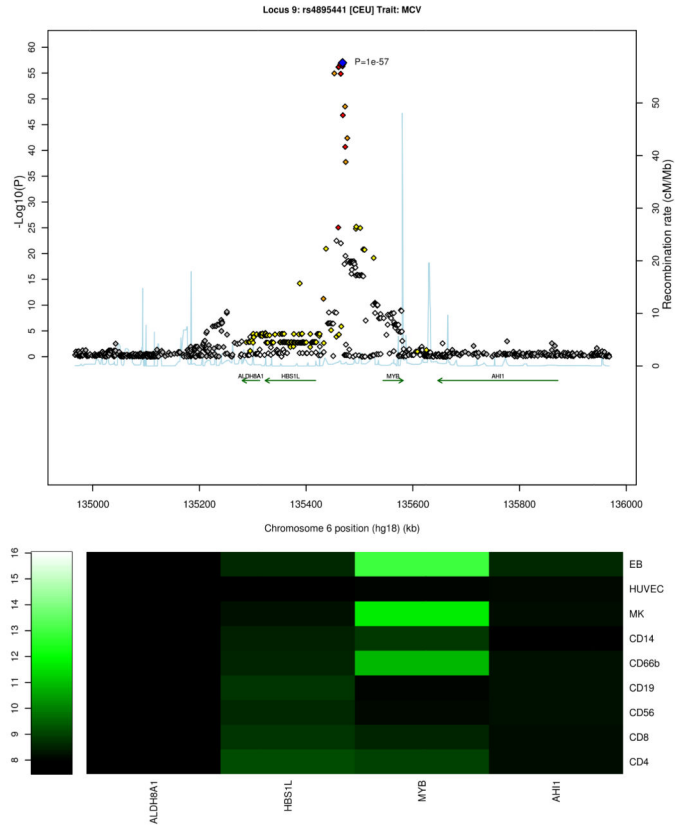


Figure 3. Gene expression in blood and endothelial cells for genes in the chromosome 6q23.3 region. (a) SNPs in the locus are plotted against recombination rates as observed in HapMap CEU, using a window of +/-500kb around the lead SNP identified in this locus, which is plotted in blue. SNPs identified by the CHARGE meta-analyses are colored according to correlation with the lead SNP ($r^2 \geq 0.8$ red; $0.5 \leq r^2 < 0.8$ orange; $0.2 \leq r^2 < 0.5$ yellow; $r^2 < 0.2$ white; no r^2 value provided). The P value for the lead SNP in this region is provided. (b) A heatmap of gene expression levels in nine blood and endothelial cell lines is shown, including all genes, as annotated by ENSEMBL 54, within the +/- 500kb window of the locus (MK = megakaryocyte; EB = erythroid bodies; HUVEC = human umbilical vein endothelial cells; CD14 = monocytes; CD66b = granulocytes; CD19 = B lymphocytes; CD56 = NK cells; CD8 = Tc lymphocytes; CD4 = Th lymphocytes).

Table 1

CHARGE cohort description

	AGES	ARIC	CHS	FHS	RS	InCHIANTI
Number of individuals eligible for GWAS	3219	8127	3275	3381	5523	1206
Percent women	58	53	61	54	60	56
Mean age (yrs)	51 (6)	54 (6)	72 (5)	38 (9)	68 (8)	68 (16)
Hb (g/dL)	13.40 (1.45)	14.80 (1.02)	14.11 (1.23)	14.46 (1.37)	14.12 (1.28)	13.77 (1.35)
Hct (%)	40.35 (3.53)	43.60 (2.86)	42.14 (3.54)	43.00 (3.9)	41.36 (3.35)	40.63 (3.49)
MCH (picogram)	30.91 (1.69)	NA	NA	30.64 (1.85)	30.16 (1.81)	3.05 (1.99)
MCHC (%)	33.57 (0.70)	NA	NA	33.67 (0.91)	34.17 (1.15)	33.84 (1.05)
MCV (femtoliter)	92.08 (4.49)	90.7 (4.22)	NA	90.7 (5.0)	88.29 (4.30)	90.22 (4.84)
RBC (1M cell/cmm)	4.39 (0.41)	NA	NA	4.73 (0.46)	4.69 (0.44)	4.51 (0.43)
Years of baseline examinations	1968-1991	1987-1989	1989-90	1971-1975	1990-1993	1998
Years of DNA collection	2002-2006	1987-1998	1989-1990	1996-1999	1990-1993	1998-2001

Sample sizes and summary statistics of covariates and erythrocyte traits measured in each cohort in CHARGE. Erythrocyte traits are abbreviated as: hemoglobin concentration (Hgb), hematocrit (Hct), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), erythrocyte count (RBC). The sample sizes presented are the range of numbers of individuals with both genotype and erythrocyte measures, after the exclusion of individuals with values beyond three standard deviations of the population mean for each erythrocyte trait.

Values for age and each erythrocyte trait are presented as mean(sd).

Table 2
CHARGE discovery meta-analysis results, ordered by genomic locus

Locus#	Chr	Hgb	Hct	MCH	MCHC	MCV	RBC	In RefGene	Gene Annotation		Closest RefGene
									RefGeneswithin60kb		
1	1q23.1	0	0	0	8	0	0	SPTA1	OR10Z1; SPTA1; OR10X1; OR6Y1		
2	2p21	3	6	0	0	0	0	PRKCE	PRKCE		
3	2p16.1	0	0	0	0	27	0	BCL11A	BCL11A		
4	3q29	0	0	4	0	11	0	TFR3	TFR3		
5	4q12	0	0	0	0	8	0	none	none	KIT	
6	6p22.2	49	3	72	0	133	0	HFE; LRRRC16; SCGN; SLC17A1; SLC17A3; SLC17A2; TRIM38; HIST1H4B; ZNF322A	HFE; SCGN; LRRRC16; SLC 17A1; SLC17A2; SLC17A3; SLC17A4; ZNF322A; ABT1; TRIM38		
7	6p21.1	0	0	51	0	65	0	PRICKLE4;FRS3;TFEB; MED20;USP49;CCND3;BYSL	PRICKLE4; FRS3; PGC; TFEB; MED20; USP49; CCND3; BYSL; TAF8; PGC; FRS3		
8	6q21	0	0	0	0	5	0	none	CD164		
9	6q23.3	0	14	43	13	83	24	HBS1L; MYB	ALDH8A1; HBS1L; MYB		
10	6q24.1	0	0	9	0	13	0	none	none	CTED2	
11	7p12.2	0	0	0	0	5	0	IKZF1	IKZF1		
12	7q22.1	0	2	0	0	2	2	TFR2; ZAN	GNB2; PCOLCE; FBXO24; TFR2; ACTL6B; HRBL; MOSPD3; LRCH4; ZAN; EPO; POP7; PERQ1; EPHB4		
13	7q36.1	1	3	0	0	0	0	PRKAG2	PRKAG2		
14	9p24.1	0	0	9	0	19	0	RCL1	RCL1; AK3		
15	10q11.21	0	0	0	0	4	0	MARCH8	MARCH8; ALOX5		
16	10q21.3	1	3	0	0	0	0	HK1	HK1		
17	12q24.12	10	9	0	0	0	0	SH2B3; ATXN2; c12orf30; PTPN11	SH2B3; ATXN2; BRAP; ACAD10; ERP29; TMEM116; C12orf30; TRAFD1; C12orf30; RPL6; PTPN11		
18	14q23.3	0	0	0	0	9	0	MAX;FNTB	MAX; RAB15; FNTB		
19	16p13.3	0	0	1	0	1	0	ITFG3	LUC7L; PDIA2; AXIN1; ITFG3; RGS11; ARHGDIG		
20	19p13.13	0	0	6	0	25	0	MAN2B1; RTBDN; MAST1; DNASE2; GCDH; FARSA	MAN2B1; MORG1; ZNF490; TNPO2; DHPS; FBXW9; ZNF791; C19orf56; JUNB; HOOK2; PRDX2; RNASEH2A; RTBDN; GADD45GIP1; KLF1;		

Locus#	Chr	Hgb	Hct	# SNPs per trait							Gene Annotation		Closest RefGene
				MCH	MCHC	MCV	RBC	In RefGene	RefGeneswithin60kb				
21	20q13.2	3	0	0	0	0	0	0	0	none	none	TSHZ2	
22	22q12.3	13	2	9	0	36	0	C22orf33; TST; MPST	MPST	IL2RB; C22orf33; KCTD17; TST; TMPRSS6; MPST			
23	22q13.33	0	0	0	0	12	0	TMEM112B; NCAPH2; SCO2; ECGF1	TMEM112B; ADM2; MIOX; ECGF1; KLHDC7B; LOC440836; SAFP2; SCO2; NCAPH2; SBF1; MAPK8IP2; MIOX; CHKB; CPT1B				

CHARGE meta-analysis results, showing the chromosomal position of each locus identified the number of SNPs identified within each and for each erythrocyte trait with meta-analysis P value $< 5 \times 10^{-8}$. Annotation for SNPs within genes (InRefGene), within ± 60 kb of annotated RefGenes (RefGenewithin60kb), or in cases where no annotated gene was identified within 60kb, the nearest gene is reported (ClosestRefGene).

Table 3
CHARGE meta-analysis results, ordered by locus and trait, and HaemGen replication analysis

Locus#	Trait	SNPID	Chr	PhysPos	min_all	MAF	Gene	CHARGE				HaemGen				CHARGE + HaemGen			
								% Var	Beta	s.e.	P	Beta	s.e.	P	Beta	s.e.	P	Beta	s.e.
2	Hct	rs10168349	2	46208555	C	0.341667	PRKCE	0.16%	0.201	0.0283	1.176E-12	0.1524	0.0471	0.001212	0.1875	0.0238	3.748E-15		
6	Hct	rs1800562	6	26201120	A	0.041667	HFE	0.09%	0.3747	0.0608	7.204E-10	0.1167	0.1004	0.245	0.3073	0.0513	2.035E-09		
9	Hct	rs9483788	6	135477194	C	0.177966	HBS1L/MYB	0.13%	0.2172	0.0328	3.551E-11	0.2147	0.0527	4.55E-05	0.2166	0.0274	2.811E-15		
12	Hct	rs7385804	7	100073906	C	0.377358	TFR2	0.17%	-0.1592	0.0286	2.745E-08	-0.1269	0.0478	0.00799	-0.151	0.0242	4.45E-10		
13	Hct	rs10224002	7	151045974	G	0.258333	PRKAG2	0.22%	0.1691	0.0299	1.492E-08	0.2727	0.0493	3.08E-08	0.1963	0.0252	6.045E-15		
16	Hct	rs16926246	10	70763398	T	0.110169	HK1	0.13%	0.337	0.0513	4.986E-11	0.3094	0.096	0.00127	0.3315	0.0445	9.636E-14		
17	Hct	rs11065987	12	110556807	G	0.341667	SH2B3/ATXN2	0.15%	-0.1809	0.0288	3.343E-10	-0.1438	0.0465	0.001983	-0.171	0.0241	1.363E-12		
22	Hct	rs2413450	22	35800170	T	0.381818	TMPRSS6	0.10%	-0.162	0.0279	6.333E-09	-0.2082	0.0467	8.13E-06	-0.1736	0.0236	1.846E-13		
2	Hgb	rs10495928	2	46206670	G	0.341667	PRKCE	0.10%	0.011	0.011	5.926E-10	0.0537	0.0152	0.000414	0.0631	0.0088	7.052E-13		
6	Hgb	rs1800562	6	26201120	A	0.041667	HFE	0.17%	0.0224	0.0224	3.592E-16	0.1167	0.0322	0.000295	0.1617	0.0182	5.737E-19		
13	Hgb	rs10224002	7	151045974	G	0.258333	PRKAG2	0.14%	0.011	0.011	4.605E-08	0.0964	0.0164	4.19E-09	0.0714	0.0091	3.025E-15		
16	Hgb	rs16926246	10	70763398	T	0.110169	HK1	0.21%	0.0192	0.0192	1.036E-09	0.0863	0.0329	0.008721	0.1099	0.0164	2.116E-11		
17	Hgb	rs11065987	12	111076069	A	0.333333	TRAFD1	0.16%	0.0107	0.0107	1.345E-10	0.0427	0.0151	0.004591	0.0585	0.0086	1.159E-11		
21	Hgb	rs6013509	20	50751758	A	0.183673	TSHZ2	0.15%	-0.0699	0.0116	1.963E-09	-0.0479	0.0206	0.01999	-0.0646	0.01	1.054E-10		
22	Hgb	rs855791	22	35792882	A	0.391667	TMPRSS6	0.20%	-0.0962	0.011	2.044E-18	-0.0845	0.0157	7.97E-08	-0.0923	0.0089	3.25E-25		
4	MCH	rs11915082	3	197293536	A	0.425	TFR3	0.24%	0.0041	0.0007	4.888E-09	0.0035	0.0009	5.66E-05	0.0038	0.0005	7.729E-13		
6	MCH	rs1408272	6	25950930	G	0.033898	SLC17A3	0.41%	-0.0134	0.0015	1.369E-18	-0.0183	0.0019	2.37E-22	-0.0153	0.0012	3.868E-39		
7	MCH	rs9349205	6	42033137	A	0.196429	CCND3/BYSL	0.29%	-0.0066	0.0008	1.785E-16	-0.0037	0.0009	2.11E-05	-0.0053	0.0006	8.198E-20		
9	MCH	rs7776054	6	135460609	G	0.220339	HBS1L/MYB	1.02%	-0.0092	0.0008	1.976E-33	-0.0107	0.0009	4.33E-36	-0.0099	0.0006	7.356E-69		
10	MCH	rs628751	6	139880112	C	0.491667	CITED2	0.34%	-0.0049	0.0007	3.84E-13	-0.0034	0.0008	7.24E-06	-0.0043	0.0005	1.262E-17		
14	MCH	rs10758658	9	4846877	A	0.186441	RCL1	0.18%	-0.0048	0.0008	1.634E-08	-0.0048	0.0009	5.14E-07	-0.0048	0.0006	2.166E-14		
19	MCH	rs1122794	16	249156	A	0.224138	ITFG3	0.28%	0.0054	0.001	2.992E-08	0.0036	0.0012	0.00299	0.0047	0.0007	2.675E-10		
20	MCH	rs11085824	19	12862547	G	0.366667	GCDH	0.20%	-0.0041	0.0007	8.105E-09	-0.003	0.0009	0.000532	-0.0037	0.0005	1.415E-11		
22	MCH	rs2413450	22	35800170	T	0.381818	TMPRSS6	0.41%	-0.006	0.0007	8.818E-17	-0.0068	0.0008	5.34E-18	-0.0064	0.0005	8.77E-34		
1	MCHC	rs857721	1	156879172	A	0.316667	SPTA1	0.33%	-0.0022	0.0004	3.414E-09	-0.0013	0.0004	0.00266	-0.0018	0.0003	1.033E-10		
9	MCHC	rs9373124	6	135464902	C	0.220339	HBS1L/MYB	0.30%	-0.0023	0.0004	6.486E-10	-0.0018	0.0004	2.6E-05	-0.0021	0.0003	7.003E-14		
3	MCV	rs2540917	2	60462263	C	0.433333	BCL11A	0.24%	-0.0031	0.0005	2.127E-11	-0.0022	0.0006	0.000192	-0.0028	0.0004	1.125E-14		

Locus#	Trait	SNPID	Chr	PhysPos	min_all	MAF	Gene	CHARGE				HaemGen				CHARGE + HaemGen			
								% Var	Beta	s.e.	P	% Var	Beta	s.e.	P	% Var	Beta	s.e.	P
4	MCV	rs9859260	3	197284944	C	0.35	TFRC	0.23%	0.003	0.0005	3.246E-10	0.003	0.0008	0.000166	0.003	0.0004	8.495E-14		
5	MCV	rs172629	4	55102519	G	0.116667	KIT	0.27%	-0.0043	0.0007	1.36E-09	-0.0051	0.001	2.22E-07	-0.0046	0.0006	9.81E-16		
6	MCV	rs1800562	6	26201120	A	0.041667	HFE	0.58%	0.0115	0.0011	1.425E-27	0.0137	0.0015	5.02E-20	0.0122	0.0009	1.012E-46		
7	MCV	rs9349205	6	42033137	A	0.196429	CCND3/BYSL	0.58%	-0.0055	0.0005	1.756E-24	-0.0043	0.0008	8.48E-08	-0.0051	0.0004	1.121E-31		
8	MCV	rs9374080	6	109723113	C	0.375	CD164	0.22%	-0.0026	0.0005	3.695E-08	-0.0017	0.0006	0.003738	-0.0023	0.0004	4.198E-10		
9	MCV	rs4895441	6	135468266	G	0.225	HBS1L/MYB	1.12%	-0.008	0.0005	1.004E-57	-0.0083	0.0008	3.03E-27	-0.0081	0.0004	7.241E-86		
10	MCV	rs643381	6	139881116	A	0.489362	CITED2	0.50%	-0.0039	0.0005	2.663E-18	-0.0037	0.0007	1.63E-07	-0.0039	0.0004	4.665E-25		
11	MCV	rs12718597	7	50395922	A	0.275	IKZF1	0.26%	0.0032	0.0005	8.138E-12	0.0018	0.0007	0.0145	0.0028	0.0004	4.689E-13		
12	MCV	rs7786877	7	100051951	G	0.208333	TFR2	0.13%	0.0032	0.0005	5.452E-09	0.0024	0.0008	0.002081	0.003	0.0004	2.543E-11		
14	MCV	rs10758658	9	4846877	A	0.186441	RCL1	0.29%	-0.0041	0.0006	4.354E-13	-0.0045	0.0008	4.65E-08	-0.0043	0.0005	3.184E-20		
15	MCV	rs11239550	10	45344735	G	0.228814	MARCH8	0.15%	-0.0028	0.0005	1.873E-08	-0.0022	0.0007	0.003496	-0.0026	0.0004	1.346E-10		
18	MCV	rs4466998	14	64545293	A	0.478723	FNTB	0.17%	0.0027	0.0005	8.925E-09	0.0008	0.0006	0.2061	0.002	0.0004	4.907E-08		
19	MCV	rs7189020	16	244804	T	0.431034	ITFG3	0.19%	-0.0031	0.0005	1.081E-09	-0.0024	0.0007	0.000871	-0.0029	0.0004	1.819E-12		
20	MCV	rs7255045	19	12793269	A	0.25	RTBDN	0.27%	-0.0037	0.0006	1.233E-11	-0.0018	0.0008	0.0266	-0.0032	0.0004	2.173E-12		
22	MCV	rs2413450	22	35800170	T	0.381818	TMPRSS6	0.65%	-0.0054	0.0005	1.078E-30	-0.0046	0.0007	7.62E-11	-0.0052	0.0004	2.772E-41		
23	MCV	rs131794	22	49318618	A	0.166667	ECGF1	0.22%	-0.0044	0.0006	2.189E-13	-0.0029	0.0009	0.001795	-0.004	0.0005	1.033E-15		
9	RBC	rs9483788	6	135461324	G	0.225	HBS1L/MYB	0.65%	0.0141	0.0016	3.115E-19	0.0155	0.0013	2.19E-30	0.0141	0.001	1.148E-47		
12	RBC	rs2075671	7	100183042	A	0.225	EPO	0.20%	0.0086	0.0016	3.058E-08	0.0047	0.0016	0.003383	0.0068	0.0011	1.123E-09		

Replication test results for the lead SNP per locus and per erythrocyte trait (45 SNPs). Results are organized by trait, the by locus, as indicated in Table 1. Results from a combined CHARGE and HaemGen Consortium meta-analysis are presented. Minor allele frequency (MAF) is presented based on HapMap CEU. % Var indicates the percent of variance explained by the lead SNP in the corresponding trait-locus. *P* values in bold font meet a Bonferroni-corrected significance threshold for replication of $P < 0.0011$ (0.05/45). Units were Hgb g/dL, Hct %, MCH picogram, MCHC g/dL, MCV femtoliter, RBC 1 M cells/ccm.