Human Molecular Genetics, 2013, Vol. 22, No. 7 **1457–1464**
doi:10.1093/hmg/dds534
Advance Access published on December 20, 2012

GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children

Jin Li[1], Joseph T. Glessner[1], Haitao Zhang[1], Cuiping Hou[1], Zhi Wei[3], Jonathan P. Bradfield[1], Frank D. Mentch[1], Yiran Guo[1], Cecilia Kim[1], Qianghua Xia[2], Rosetta M. Chiavacci[1], Kelly A. Thomas[1], Haijun Qiu[1], Struan F.A. Grant[1,2,4], Susan L. Furth[4], Hakon Hakonarson[1,2,4] and Patrick M.A. Sleiman[1,2,4,*]

[1]Center for Applied Genomics, Abramson Research Center and [2]Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA, [3]Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA, [4]Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

Received August 16, 2012; Revised October 26, 2012; Accepted December 13, 2012

Hematological traits are important clinical indicators, the genetic determinants of which have not been fully investigated. Common measures of hematological traits include red blood cell (RBC) count, hemoglobin concentration (HGB), hematocrit (HCT), mean corpuscular hemoglobin (MCH), MCH concentration (MCHC), mean corpuscular volume (MCV), platelet count (PLT) and white blood cell (WBC) count. We carried out a genome-wide association study of the eight common hematological traits among 7943 African-American children and 6234 Caucasian children. In African Americans, we report five novel associations of *HBE1* variants with HCT and MCHC, the alpha-globin gene cluster variants with RBC and MCHC, and a variant at the *ARHGEF3* locus with PLT, as well as replication of four previously reported loci at genome-wide significance. In Caucasians, we report a novel association of variants at the *COPZ1* locus with PLT as well as replication of four previously reported loci at genome-wide significance. Extended analysis of an association observed between MCH and the alpha-globin gene cluster variants demonstrated independent effects and epistatic interaction at the locus, impacting the risk of iron deficiency anemia in African Americans with specific genotype states. In summary, we extend the understanding of genetic variants underlying hematological traits based on analyses in African-American children.

INTRODUCTION

Disorders of the hematopoietic system are associated with a variety of diseases. Several studies have now been reported on the genetic determinants of blood cell traits, primarily in adult populations of European ancestry (1–9) or East Asian ancestry (10–12). Combined, these studies have identified more than 100 loci associated with blood cell quantitative traits (13). Among common measures were white blood cell (WBC), red blood cell (RBC), hemoglobin concentration (HGB), hematocrit (HCT), mean corpuscular hemoglobin (MCH), MCH concentration (MCHC), mean corpuscular volume (MCV) and platelet count (PLT).

In African Americans, WBC and neutrophil counts have been associated with the Duffy antigen receptor for chemokines (*DARC*) locus by admixture mapping (14,15). Two genome-wide association (GWA) studies have been reported on WBC (16) and PLT phenotypes (17). The WBC genome-wide association study (GWAS) reported association at a locus on 4q13 (16), whereas the PLT GWAS uncovered 10 PLT associated loci and three loci for the mean platelet volume (MPV) (17). Finally, in a custom chip study by Lo

*To whom correspondence should be addressed at: Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA. Tel: +1 2674267653; Fax: +1 2674260363; Email: sleimanp@email.chop.edu

Table 1. Genetic variants associated with hematological traits in African-American children

Trait	SNP	Chr	Position (hg18)	Gene	Minor/ major allele	MAF	β	SE	*P*(AA)	Ref	Additional number of SNPs[a]
HCT	rs2213169	11	52 596 39	HBE1/HBB/HBD/ HBBP1/HBG1 cluster	T/C	0.1362	−0.4471	0.06792	4.94×10^{-11}		2
HGB	rs7203560	16	124 390	16p13.3	G/T	0.06081	−0.199	0.03555	2.23×10^{-8}	18	0
RBC	rs7203560	16	124 390	16p13.3	G/T	0.06081	0.1452	0.01451	2.01×10^{-23}		1
	rs5987027	X	153 667 301	MPP1	T/C	0.2462	−0.07447	0.01095	1.199×10^{-11}	18	0
MCH	rs7203560	16	124 390	16p13.3	G/T	0.06081	−1.255	0.07487	5.18×10^{-62}	18	37
MCHC	rs7203560	16	124 390	16p13.3	G/T	0.06081	−0.5311	0.04375	1.31×10^{-33}		8
	rs2213169	11	52 596 39	HBE1/HBB/HBD/ HBBP1/HBG1 cluster	T/C	0.1362	0.2227	0.02997	1.21×10^{-13}		2
MCV	rs7203560	16	124 390	16p13.3	G/T	0.06081	−2.555	0.1851	7.79×10^{-43}	18	20
	rs5987027	X	153 667 301	MPP1	T/C	0.2462	0.9159	0.1458	3.679×10^{-10}	18	0
PLT	rs1354034	3	568 247 89	ARHGEF3	C/T	0.27	11.44	1.599	9.32×10^{-13}		0
	rs210134	6	33 648 187	BAK1	A/G	0.2675	−8.923	1.582	1.78×10^{-8}	18	0
WBC	rs4657616	1	157 237 710	DARC	G/A	0.09945	0.06124	0.004225	5.48×10^{-47}	14	401

[a]Additional number of genome-wide significant SNPs.

et al., including just over 7000 African-American samples typed on the Illumina IBC iSelect array that includes 49K single nucleotide polymorphisms (SNPs), association of an SNP (rs1050828) located in the canonical glucose-6 phosphate dehydrogenase (*G6PD*) gene was observed with multiple RBC traits in African Americans which had not been previously identified in Caucasians (18).

Here, we report a GWAS in our pediatric population of 7943 African-American children and 6234 Caucasian children. We examined the above eight common hematological traits and further meta-analyzed the results from the two ancestry specific analyses. We identified five novel associations in African Americans and one novel association in Caucasian children and further revealed independent effects and epistatic interaction at the alpha-globin gene cluster, the specific genotype states of which influence the risk of iron deficiency anemia in African Americans.

RESULTS

We conducted a GWA study in our pediatric population of 7943 African-American children and 6234 Caucasian children, examining the eight common hematological traits, WBC, RBC, HGB, HCT, MCH, MCHC, MCV and PLT. The characteristics of each blood cell trait and other phenotypes are summarized in Supplementary Material, Table S1. We further meta-analysed the results from the two ancestry specific analyses. Quantile–quantile plots for each analysis are shown in Supplementary Material, Figure S2. The genomic inflation factor for each analysis was 1, indicating that there was no substantial stratification in the analysis.

A total of 15 SNPs at five loci reached genome-wide significance for the eight blood trait phenotypes among Caucasian children (Supplementary Material, Fig. S3), including a novel association between variants in *COPZ1* gene and PLT (rs4326844, $P = 4.57 \times 10^{-8}$). The remaining four genome-wide significant loci were replications of previously reported associations; two loci with MCH and MCV: *TMPRSS6* (rs855791, $P = 5.32 \times 10^{-14}$ for MCH; and rs855791, $P = 2.03 \times 10^{-9}$ for MCV) and *HBS1L-MYB* (rs7775698, $P = 3.98 \times 10^{-13}$ for MCH; and rs7775698, $P = 1.55 \times 10^{-9}$ for MCV); one locus with PLT: *ARHGEF3* (rs1354034, $P = 4.35 \times 10^{-9}$) and one locus with WBC (rs389884, $P = 2.09 \times 10^{-8}$) (Supplementary Material, Table S3).

In the African-American cohort, 447 SNPs at six loci surpassed genome-wide significance (Supplementary Material, Fig. S3). The large SNP count was inflated by the association at the DARC locus with WBC count. We report novel genome-wide significant associations between variants at the epsilon-globin gene cluster and HCT (rs2213169, $P = 4.94 \times 10^{-11}$) and MCHC (rs2213169, $P = 1.21 \times 10^{-13}$) and the alpha-globin gene cluster with both RBC (rs7203560, $P = 2.01 \times 10^{-23}$) and MCHC (rs7203560, $P = 1.31 \times 10^{-33}$) and *ARHGEF3* variants (Table 1), which had previously been associated with MPV in Caucasians (4,6), with PLT in African Americans (rs1354034, $P = 9.32 \times 10^{-13}$). We also replicated at genome-wide significance, the previously reported associations between variants at the alpha-globin gene cluster and HGB, MCH and MCV (18) (Table 1), the association between the *DARC* locus and the WBC (14) as well as the *BAK1* gene and PLT (18). Restricting our analysis to African-American females, we also found that SNP, rs5987027, on the X chromosome was associated with RBC and MCV (Table 1), thus replicating the *G6PD* association reported by Lo *et al.* (18).

In addition to the results of the individual ancestry-specific cohorts described above, in a meta-analysis of both cohorts ($n = 14\,177$), variants at the 6p22.2 (*HFE*) and 6q24.1 (*CITED2*) loci that were previously reported in the CHARGE consortium study (1) reached genome-wide significance for association with MCH. The 6q24.1 locus was also genome-wide significant for MCV (Supplementary Material, Table S4). Further, three SNPs at the *ATP2B4* gene approached, but did not surpass the genome-wide significance level for association with MCHC in the meta-analysis (rs1541252, $P = 8.892 \times 10^{-8}$; rs1419114, $P = 9.887 \times 10^{-8}$; rs10900588, $P = 1.04 \times 10^{-7}$). The three SNPs showed association in both ancestry-specific GWAS with *P*-values of 10^{-5} and 10^{-4} in the African-American cohort

Table 2. Conditional analysis for variants associated with the MCH trait at the 16p13.3 locus

SNP	Chr	Position (hg18)	Minor/major allele	MAF	β	SE	P
rs1211375	16	180 281	A/C	0.2722	−0.8062	0.08801	1.53×10^{-19}
rs7203560	16	124 390	G/T	0.08869	−1.142	0.139	4.22×10^{-16}
rs7200589	16	289 332	A/G	0.1113	−1.033	0.1299	3.46×10^{-15}
rs1203957	16	181 211	T/G	0.3426	0.6516	0.08415	1.71×10^{-14}
rs11248914	16	233 563	C/T	0.3971	0.5447	0.08067	2.03×10^{-11}
rs17136255	16	340 476	T/C	0.2112	−0.598	0.09585	5.63×10^{-10}
rs2562182	16	73 946	T/C	0.2356	0.5609	0.09428	3.31×10^{-9}
rs1203981	16	205 160	C/T	0.1966	−0.5858	0.09943	4.68×10^{-9}
rs6600191/rs1211375	16	235 796	C/T	0.253	0.5028	0.09059	3.33×10^{-8}
rs7203560/rs1211375	16	124 390	G/T	0.08869	−0.7881	0.1466	8.73×10^{-8}
rs7200589/rs1211375	16	289 332	A/G	0.1113	−0.7248	0.135	9.11×10^{-8}
rs1203981/rs1211375,rs6600191	16	205 160	C/T	0.1966	−0.557	0.1006	3.64×10^{-8}
rs7203560/rs1211375,rs6600191,rs1203981	16	124 390	G/T	0.08869	−0.8147	0.1489	5.14×10^{-8}

Table 3. Interactions between SNPs at the 16p13.3 locus for RBC traits

Phenotype	Interacting SNPs SNP1	SNP2	AA (subset)[a] BETA (interaction)	P	AA[b] BETA (interaction)	P	Adjusted BETA[c] (interaction)	Adjusted P[c]
MCH	rs1211375	rs1203981	−0.7463	7.41×10^{-7}	−0.7488	7.124×10^{-30}	−0.5383	3.32×10^{-14}
	rs7203560	rs6600191	0.7323	0.00716	0.4564	0.001015	0.2775	0.04585
HCT	rs1211375	rs1203981	−0.05289	0.7787	−0.2218	0.01444	−0.03295	0.7188
HGB	rs1211375	rs1203981	−0.1618	0.01865	−0.1916	4.731×10^{-9}	−0.1117	7.58×10^{-4}
RBC	rs1211375	rs1203981	0.07872	0.003756	0.06902	1.244×10^{-8}	0.06007	1.15×10^{-5}
MCHC	rs1211375	rs1203981	−0.3908	2.095×10^{-6}	−0.2824	1.432×10^{-14}	−0.2335	1.41×10^{-8}
MCV	rs1211375	rs1203981	−1.425	0.0001849	−1.607	1.303×10^{-21}	−1.122	2.05×10^{-10}

[a]A subset of African-American children who have no Caucasian local ancestry at this region.
[b]All African-American children in this study.
[c]BETA and P-value in the linear regression models adjusting for local ancestry and other covariates.

and Caucasian cohort, respectively, with consistent direction of effect yet the final P-value decreased on meta-analysis. In total, we replicated 95 out of the 107 previously reported loci at nominal significance (Supplementary Material, Table S5).

The association at the alpha-globin gene clusters and MCH in African Americans spans 1.458 Mb and contains multiple genome-wide associated SNPs. We carried out conditional regression analyses to identify potential independent signals within the locus. To avoid confounding of the conditional regression due to the admixture in the African-American individuals, we first inferred local ancestry using the HAPMIX (19) package and selected the subset of 1642 individuals with only African ancestry in the region (hg18 chr16:37354-651906). Within the subset of samples, rs1211375 showed the most significant association ($P = 1.53 \times 10^{-19}$) (Table 2) with MCH. Following conditional analysis on rs1211375, three SNPs in this region still showed significant or borderline significant association with MCH (Table 2), indicating the presence of at least one independent signal. In a second round of conditional analyses on rs1211375 and rs6600191, all of the primary associations were ablated; however, genome-wide association at rs1203981 was restored. Similarly, association at rs7203560 was restored after conditioning on rs1211375, rs6600191 and rs1203981 (summarized in Table 2). These results suggest the presence of at least two independent signals associating with MCH at the 16p13.3 locus as well as epistatic interactions between some of the variants. To estimate the population attributable risk of these variants on the MCH phenotype, we carried out a mixed linear model analysis of variance explained by these SNPs. The proportion of phenotypic variance explained by rs1211375 and the seven other genome-wide significant SNPs in LD with it was estimated at 5.45%, while the addition of rs6600191, the independent signal at the locus, increased the proportion of phenotypic variance explained to 7.38%. We were unable to test the effects of these SNPs in African-American individuals which were homozygous for European ancestry alleles at this locus, as there were no such individuals in our study.

To further explore the potential epistatic interactions at the 16p13.3 locus, we carried out pairwise epistasis tests between the top significant SNPs of each conditional test as detailed in Table 2 against all other SNPs at the locus. We first tested for epistasis amongst the subset of African-American children who have no Caucasian local ancestry at this locus. Two pairwise interactions surpassed significance threshold ($P = 0.0083$) adjusting for multiple tests, the first between SNPs rs7203560 and rs6600191 with a P-value of 0.00716; the other between SNPs rs1211375 and rs1203981 with a P-value of 7.41×10^{-7} (Table 3). We subsequently expanded the analysis to the entire African-American cohort, which yielded P-values of 0.001015 and 7.12×10^{-30}, respectively, confirming the significance of these interactions. The interaction between rs1211375 and rs1203981 reached well beyond genome-wide significance ($P = 4.0 \times 10^{-13}$) for pairwise epistatic interactions. In addition to the MCH phenotype, significant interactions between rs1211375 and rs1203981 were also present for the other RBC phenotypes (Table 3). This pair of interaction also reached genome-wide significance with P-values of 1.43×10^{-14} and 1.30×10^{-21} for MCHC and MCV, respectively, in the entire African-American cohort. Controlling for local ancestry and all other covariates, as specified above, in the larger African-American cohort the interactions between rs1211375 and rs1203981 remained significant for all the RBC phenotypes except HCT. Furthermore, this pairwise interaction still surpassed genome-wide significance for phenotype MCH ($P = 3.32 \times 10^{-14}$) (Table 3). Imputing genotypes at this locus up to the density of the 1000 Genomes data did not reveal any additional independent signals or novel epistatic interactions; however, the interaction between rs1211375 and rs1203981 was confirmed in the imputed data (Supplementary Material, Table S6).

We further examined the effect of interaction of SNPs, rs1211375 and rs1203981, on the MCH phenotype. We stratified the sample set by rs1211375 genotype and plotted the MCH value for each rs1203981 genotype (Fig. 1C). The presence of the minor C allele of the rs1203981 SNP on a background of the rs1211375 minor allele A was associated with a significant additive decrease in the MCH value. The mean MCH level for individuals with an AA/CC genotype at rs1211375/rs1203981 was 23.42 (SD = 2.34), compared with the population mean of 27.19 placing within the range of iron deficiency anemia (http://www.compsim.com/demos/d60/Anemia.htm, date last accessed 21 December 2012). Returning to the medical records, we confirmed an enrichment of iron deficiency anemia in children with an AA/CC genotype, 20% versus 8.24%. Comparison using Fisher's exact test yielded a two-side P-value of 0.023, suggesting that African Americans with an rs1211375 AA and rs1203981 CC genotype are at higher risk of developing iron deficiency anemia than the rest of African-American population. We also found that the group of the children with an rs1211375 AA and rs1203981 CC genotype had a lower median value of HCT, HGB, MCHC, MCV and a higher median value of RBC, compared with other genotype groups. Statistical analysis indicated that among African-American children with rs1211375 AA genotype, rs1203981 showed nominal significant association with each of the RBC phenotypes (Fig. 1D and Supplementary Material, Table S7).

DISCUSSION

To the best of our knowledge, this is the first study to have examined associations between multiple hematological traits and SNP genotypes in a pediatric cohort. We assessed eight common hematological traits in two pediatric cohorts of different ancestries. Among the previously reported loci, 95 were replicated at nominal significance; four loci reached genome-wide significance in the African-American cohort and four loci reached genome-wide significance in the Caucasian cohort. We also found five novel associations that have not been previously reported in African Americans and an additional novel association among Caucasian children. Finally, we identified independent signals and epistatic interactions at the alpha-globin cluster, which impact the risk of iron deficiency anemia in African Americans.

With several large studies now reported for the same hematological traits in cohorts of different ancestries, it is possible to begin to compare the underlying genetic architectures between these populations. Comparing our results with those previously published in populations of European and East Asian ancestries highlights loci that show differential association with the traits between the different ancestral groups. While most loci show consistent patterns of association across ethnicities others such as the epsilon-globin gene cluster on chromosome 11p15.4 show striking differences. In African Americans, the locus is highly associated with HCT and MCHC; however, neither our study nor the two recent large-scale GWAS by the HaemGen (6) consortium and the CHARGE consortium (1) report association of this locus in subjects of European ancestry.

In addition to such categorical differences, qualitative differences in the effect sizes and span of association between the different ancestral groups are also evident. The alpha-globin cluster on 16p13.3 is associated with MCV and MCH in both Caucasians and African Americans; however, the effect sizes are larger in the African Americans and the associated region extends over 1.458 Mb, similarly to the extended association of the DARC locus with WBC traits, which may reflect selective pressure during evolution.

Our study also adds to the increasing number of pleiotropic loci underlying the hematological traits. Among African Americans, the epsilon-globin gene cluster is significantly associated with HCT and MCHC and the alpha-globin gene cluster is associated with HGB, RBC, MCH, MCHC and MCV. One consistent trend is for higher correlation between the RBC indices, or ratios: MCH (HGB/RBC), MCV (HCT/RBC) and MCHC (MCH/MCV) than for the absolute counts hemoglobin concentration (HGB), hematocrit (HCT) and RBC count. Such phenomena have been generally observed among other ethnic groups too (1,6,10). Okada and Kamatani attributed this finding to a difference of robustness of confounding factors (13).

Finally, we identified independent signals in the region of alpha-globin gene cluster, demonstrating the presence of allelic heterogeneity at the locus in African Americans, and identified a novel epistatic interaction that increases the risk of iron deficiency anemia. As discussed above, the two SNPs that show independent effects and epistatsis in the African Americans are not associated with RBC traits in Caucasians which would preclude this discovery being made in the Caucasian cohorts published to date. The study by Lo *et al.* (18) in which the associations between this locus and HGB and MCV were first reported in African Americans was conducted on the iSELECT array which does not include rs1203981 one of the SNPs that we now show interact.

In iron deficiency anemia, a lack of iron results in a reduced number of RBCs. Both environmental factors and genetic factors can cause iron deficiency anemia. Several loci have been identified to be associated with serum iron concentration or markers of serum iron levels (20–22), such as genes

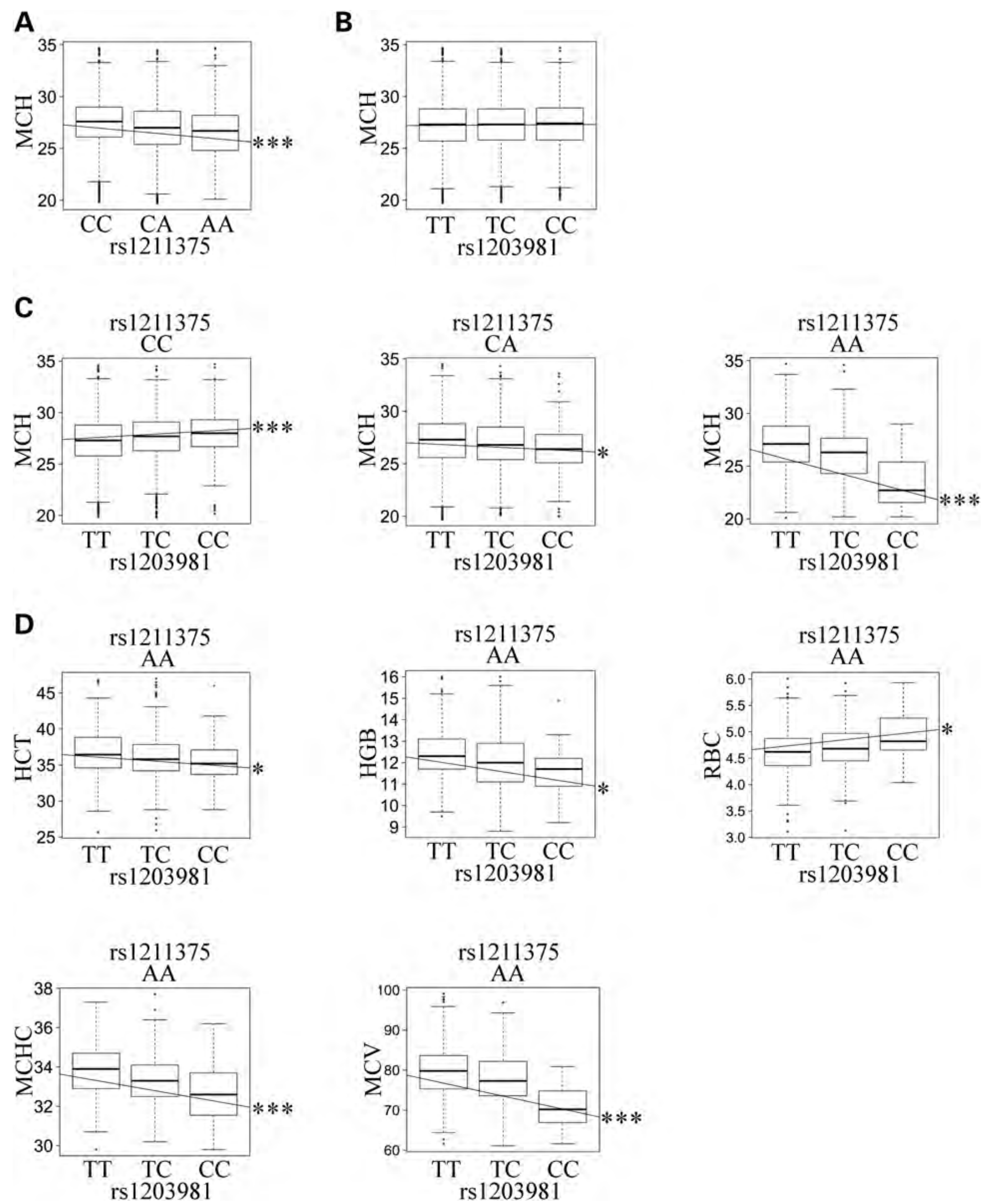

Figure 1. Box plot showing the distribution of the RBC traits in each of the genotype groups and the corresponding linear regression line. The *x*-axis shows the genotypes at each SNP indicated, and the *y*-axis shows the level of the RBC trait. (**A**) Box plot for the MCH level and genotype at rs1211375; (**B**) Box plot for MCH level and genotype at rs1203981; (**C**) Box plot for the MCH level and genotype combinations at rs1211375 and rs1203981; (**D**) Box plot for RBC traits HCT, HGB, MCHC, MCV, RBC and genotype combinations at rs1211375 and rs1203981. $^{*}P < 0.05$, $^{***}P < 5 \times 10^{-8}$; linear regression.

HFE (hemochromatosis), *TMPRSS6* (transmembrane serine protease 6), *TF* (transferrin) and *TFR2* (transferrin receptor 2). Several genes have also been related to iron deficiency anemia. For example, polymorphism GPIa-C807T in platelet collagen receptor GPIaIIa was found to affect iron deficiency anemia in young women (23). Mutations in the *TMPRSS6* gene have also been reported to cause iron-refractory iron deficiency anemia (24–27). In our study, the two interacting SNPs are both located in introns of gene *LUC7L* flanking the alpha-globin gene cluster with 25 kb distance in between. Missing or defective alpha-globin genes can cause alpha thalassemia trait which phenotypically is similar to iron deficiency anemia (28). However, iron supplement is not useful to relieve the anemia clinical phenotype (http://labtestsonline.org/understanding/conditions/thalassemia/start/1, date last accessed 21 December 2012). In conclusion, we demonstrate for the first time that epistatic interactions of variants flanking the alpha-globin gene cluster are associated with increased risk of iron deficiency anemia in African Americans.

MATERIALS AND METHODS

Ethics statement

The study was approved by the Institutional Review Board at the Children's Hospital of Philadelphia, and written informed consent for sample collection and DNA genotyping was provided by the parents of all participating children.

Sample description

A total of 17 324 children were recruited to the study. Blood samples were taken and each phenotype was measured at the Children's Hospital of Philadelphia. Blood-related diseases were diagnosed according to the standard criteria. Only genetically inferred Caucasian and African-American children were included in the analyses and subjects with missing data or measurement beyond 3SD of the mean were excluded from the study for the particular trait examined. The final analyses encompassed 7943 African-American children and 6234 Caucasian children.

Iron deficiency anemia was defined in our pediatric sample cohort using the following criteria: (1) hemoglobin level $<$110 mg/ml of whole blood, and hematocrit of $<$33%; (2) altered RBC distribution width and/or reduced MCV and/or MCHC; (3) decrease in serum ferritin levels if available; (4) increase in serum-transferring levels and total iron binding capacity if available and (5) evidence of small pale RBCs on a smear, consistent with hypochronic microcytic anemia. In addition to these criteria, we were able to exclude alpha thalassemia as a cause of the anemia using the results of a neonatal genetic screen of the alpha and beta globin genes that are present in the electronic medical record.

SNP genotyping and quality control

We performed SNP genotyping of 4497 African-American samples and 3292 Caucasian samples on the Illumina 550 k chip with the remaining samples genotyped on the Human610-Quad version 1 array. After data normalization and canonical genotype clustering according to Illumina standard protocols, we only included samples with a call rate of $>$98% for further analysis. For genotyping markers, we only included those SNPs that were common to both 550 k chip and 610-Quad chip ($n = 544\,917$) and with a genotype missing rate of $<$ 5%, minor allele frequency $>$0.01, as well as Hardy-Weinberg equilibrium $P > 0.0001$. We used multi-dimensional scaling, as implemented in PLINK (version 1.06) (29), for inferring population structure in the cohort. Follow-up association analyses were performed within each ethnicity group separately. We detected cryptic relatedness between samples which have identity-by-descent (IBD) scores $>$ 0.25 among Caucasians and that $>$0.50 among African Americans, and then removed one sample from each pair. We also conducted principal component analysis among African Americans and Caucasians separately using EIGENSTRAT (30) to further quantify population relationships among these samples.

Statistical analyses

As all the blood phenotype traits of interest were approximately normally distributed (Supplementary Material, Fig. S1), we used a linear regression model to evaluate the genetic association between SNP genotypes and the quantitative blood traits. Age, sex and hematological disease status (Supplementary Material, Table S2) were incorporated in the analysis as covariates. The first three principal components from the EIGENSTRAT (30) analysis were also included as covariates in the association analysis of the African-American children and in the analysis of Caucasian children for MCH, MCV and RBC to control for population stratification. Software PLINK (29) was used for all the aforementioned regression analyses. Finally, we performed a meta-analysis of both cohorts by combining the results from each group through an inverse variance-based analytical strategy implemented in the software METAL (31) with the inclusion of heterogeneity analysis.

Local ancestry estimation

HAPMIX (19) software was used to infer local ancestry of each African-American participant, which was defined as 0, 1 or 2 European chromosomes. Phased CEU [CEPH (Utah residents with ancestry from northern and western Europe)] and YRI haplotypes from HapMap3 were used as reference panels in local ancestry estimation. A total of 1642 African-American individuals with no CEU ancestry in the region of chr16:37354-651906(hg18) were selected for further analysis.

Conditional analysis and epistasis tests

Association analysis for blood trait MCH was performed among African-American children without any CEU ancestry at the locus of alpha-globin gene cluster and a similar association analysis was performed, conditional on the topmost significant SNP, via software PLINK (29). Then additional rounds of analyses were conducted, conditional on topmost significant SNPs arisen from preceding analyses. Pairwise epistasis tests between the top significant SNPs of each conditional analysis were conducted via PLINK (29) and software R when adjusting for covariates including local ancestry, age, sex, hematological disease status and the first three principal components from the EIGENSTRAT analysis. The proportion of phenotypic variance explained by the significant SNPs was estimated using the GCTA software package (32).

SNP imputation

Imputation of the alpha-globin gene locus was performed using the IMPUTE2 package (33,34). The reference panel was the 1000 Genome Phase I integrated variant set (http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html, date last accessed 21 December 2012). Association analysis of the imputed genotypes, taking the imputation uncertainty into account, was conducted using the SNPTEST v2 package (34) missing data-likelihood score test.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank all the children who donated blood samples to the Children's Hospital of Philadelphia for genetic research purpose.

Conflict of Interest statement. None declared.

FUNDING

This project was funded by the Institute Development Funds to the Center for Applied Genomics and an Adele S. and Daniel S. Kubert Estate gift to the Center for Applied Genomics.

REFERENCES

1. Ganesh, S.K., Zakai, N.A., van Rooij, F.J., Soranzo, N., Smith, A.V., Nalls, M.A., Chen, M.H., Kottgen, A., Glazer, N.L., Dehghan, A. *et al.* (2009) Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat. Genet.*, **41**, 1191–1198.
2. Kullo, I.J., Ding, K., Jouni, H., Smith, C.Y. and Chute, C.G. (2010) A genome-wide association study of red blood cell traits using the electronic medical record. *PLoS One*, **5**, e13011.
3. Gudbjartsson, D.F., Bjornsdottir, U.S., Halapi, E., Helgadottir, A., Sulem, P., Jonsdottir, G.M., Thorleifsson, G., Helgadottir, H., Steinthorsdottir, V., Stefansson, H. *et al.* (2009) Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.*, **41**, 342–347.
4. Meisinger, C., Prokisch, H., Gieger, C., Soranzo, N., Mehta, D., Rosskopf, D., Lichtner, P., Klopp, N., Stephens, J., Watkins, N.A. *et al.* (2009) A genome-wide association study identifies three loci associated with mean platelet volume. *Am. J. Hum. Genet.*, **84**, 66–71.
5. Soranzo, N., Rendon, A., Gieger, C., Jones, C.I., Watkins, N.A., Menzel, S., Doring, A., Stephens, J., Prokisch, H., Erber, W. *et al.* (2009) A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function. *Blood*, **113**, 3831–3837.
6. Soranzo, N., Spector, T.D., Mangino, M., Kuhnel, B., Rendon, A., Teumer, A., Willenborg, C., Wright, B., Chen, L., Li, M. *et al.* (2009) A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.*, **41**, 1182–1190.
7. Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V.G., Chen, W., Usala, G., Busonero, F., Maschio, A., Albai, G. *et al.* (2008) Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc. Natl Acad. Sci. USA*, **105**, 1620–1625.
8. Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A.H., Labrune, Y. *et al.* (2011) New gene functions in megakaryopoiesis and platelet formation. *Nature*, **480**, 201–208.
9. Nalls, M.A., Couper, D.J., Tanaka, T., van Rooij, F.J., Chen, M.H., Smith, A.V., Toniolo, D., Zakai, N.A., Yang, Q., Greinacher, A. *et al.* (2011) Multiple loci are associated with white blood cell phenotypes. *PLoS Genet.*, **7**, e1002113.
10. Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y. and Kamatani, N. (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.*, **42**, 210–215.
11. Okada, Y., Kamatani, Y., Takahashi, A., Matsuda, K., Hosono, N., Ohmiya, H., Daigo, Y., Yamamoto, K., Kubo, M., Nakamura, Y. *et al.* (2010) Common variations in PSMD3-CSF3 and PLCB4 are associated with neutrophil count. *Hum. Mol. Genet.*, **19**, 2079–2085.
12. Okada, Y., Hirota, T., Kamatani, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Higasa, K., Yamaguchi-Kabata, Y., Hosono, N., Nalls, M.A. *et al.* (2011) Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS Genet.*, **7**, e1002067.
13. Okada, Y. and Kamatani, Y. (2012) Common genetic factors for hematological traits in Humans. *J. Hum. Genet.*, **57**, 161–169.
14. Nalls, M.A., Wilson, J.G., Patterson, N.J., Tandon, A., Zmuda, J.M., Huntsman, S., Garcia, M., Hu, D., Li, R., Beamer, B.A. *et al.* (2008) Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the health ABC and Jackson heart studies. *Am. J. Hum. Genet.*, **82**, 81–87.
15. Reich, D., Nalls, M.A., Kao, W.H., Akylbekova, E.L., Tandon, A., Patterson, N., Mullikin, J., Hsueh, W.C., Cheng, C.Y., Coresh, J. *et al.* (2009) Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet.*, **5**, e1000360.
16. Reiner, A.P., Lettre, G., Nalls, M.A., Ganesh, S.K., Mathias, R., Austin, M.A., Dean, E., Arepalli, S., Britton, A., Chen, Z. *et al.* (2011) Genome-wide association study of white blood cell count in 16,388 African Americans: the Continental Origins and Genetic Epidemiology Network (COGENT). *PLoS Genet.*, **7**, e1002108.
17. Qayyum, R., Snively, B.M., Ziv, E., Nalls, M.A., Liu, Y., Tang, W., Yanek, L.R., Lange, L., Evans, M.K., Ganesh, S. *et al.* (2012) A meta-analysis and genome-wide association study of platelet count and mean platelet volume in African Americans. *PLoS Genet.*, **8**, e1002491.
18. Lo, K.S., Wilson, J.G., Lange, L.A., Folsom, A.R., Galarneau, G., Ganesh, S.K., Grant, S.F., Keating, B.J., McCarroll, S.A., Mohler, E.R. III *et al.* (2011) Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum. Genet.*, **129**, 307–317.
19. Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D. and Myers, S. (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.*, **5**, e1000519.
20. McLaren, C.E., Garner, C.P., Constantine, C.C., McLachlan, S., Vulpe, C.D., Snively, B.M., Gordeuk, V.R., Nickerson, D.A., Cook, J.D., Leiendecker-Foster, C. *et al.* (2011) Genome-wide association study identifies genetic loci associated with iron deficiency. *PLoS One*, **6**, e17390.
21. Pichler, I., Minelli, C., Sanna, S., Tanaka, T., Schwienbacher, C., Naitza, S., Porcu, E., Pattaro, C., Busonero, F., Zanon, A. *et al.* (2011) Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels. *Hum. Mol. Genet.*, **20**, 1232–1240.
22. Tanaka, T., Roy, C.N., Yao, W., Matteini, A., Semba, R.D., Arking, D., Walston, J.D., Fried, L.P., Singleton, A., Guralnik, J. *et al.* (2010) A genome-wide association analysis of serum iron concentrations. *Blood*, **115**, 94–96.
23. Carlsson, L.E., Hempel, S. and Greinacher, A. (2002) Iron deficiency anaemia in young women. *Eur. J. Haematol.*, **68**, 341–344.
24. Finberg, K.E., Heeney, M.M., Campagna, D.R., Aydinok, Y., Pearson, H.A., Hartman, K.R., Mayo, M.M., Samuel, S.M., Strouse, J.J., Markianos, K. *et al.* (2008) Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA). *Nat. Genet.*, **40**, 569–571.
25. Ramsay, A.J., Quesada, V., Sanchez, M., Garabaya, C., Sarda, M.P., Baiget, M., Remacha, A., Velasco, G. and Lopez-Otin, C. (2009) Matriptase-2 mutations in iron-refractory iron deficiency anemia patients provide new insights into protease activation mechanisms. *Hum. Mol. Genet.*, **18**, 3673–3683.
26. Altamura, S., D'Alessio, F., Selle, B. and Muckenthaler, M.U. (2010) A novel TMPRSS6 mutation that prevents protease auto-activation causes IRIDA. *Biochem. J.*, **431**, 363–371.
27. De Falco, L., Totaro, F., Nai, A., Pagani, A., Girelli, D., Silvestri, L., Piscopo, C., Campostrini, N., Dufour, C., Al Manjomi, F. *et al.* (2010) Novel TMPRSS6 mutations associated with iron-refractory iron deficiency anemia (IRIDA). *Hum. Mutat.*, **31**, E1390–E1405.
28. Galanello, R. and Cao, A. (1993) Alpha-Thalassemia. In: Pagon, R.A., Bird, T.D., Dolan, C.R., Stephens, K. and Adam, M.P. (eds), *GeneReviewsTM [Internet]*. Seattle (WA): University of Washington. http://www.ncbi.nlm.nih.gov/books/NBK1116/
29. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
30. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.

31. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
32. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
33. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
34. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.