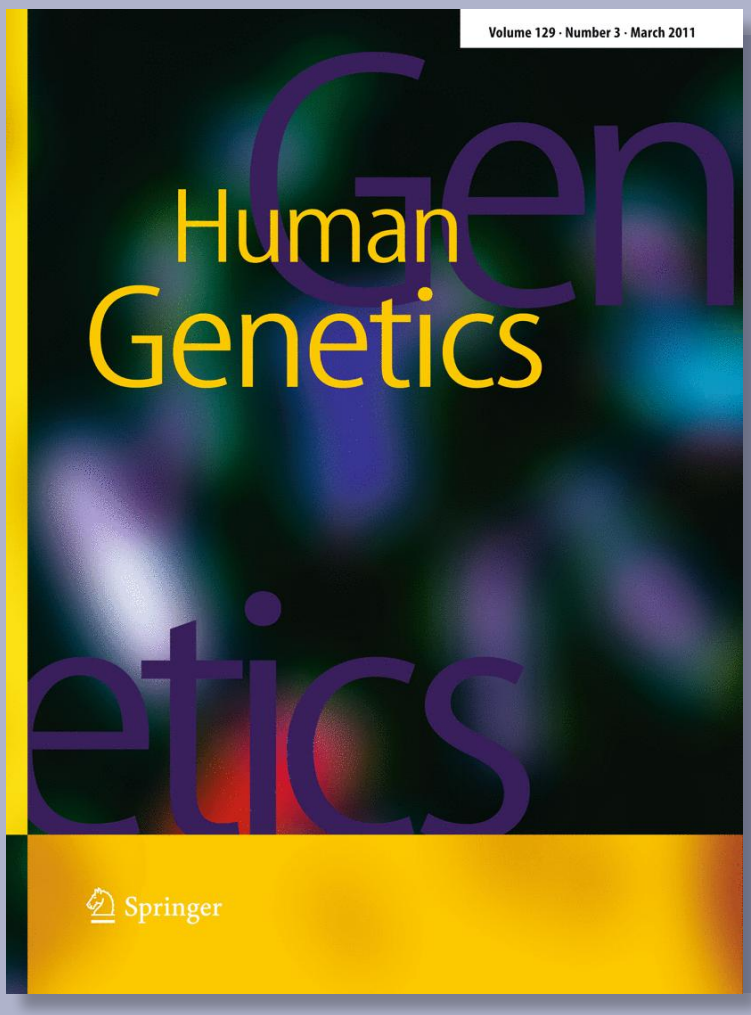
Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans

Human Genetics

ISSN 0340-6717
Volume 129
Number 3

Hum Genet (2010)
129:307-317
DOI 10.1007/
s00439-010-0925-1



Springer

Hum Genet (2011) 129:307–317
DOI 10.1007/s00439-010-0925-1

ORIGINAL INVESTIGATION

Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans

Ken Sin Lo · James G. Wilson · Leslie A. Lange · Aaron R. Folsom · Geneviève Galarneau · Santhi K. Ganesh · Struan F. A. Grant · Brendan J. Keating · Steven A. McCarroll · Emile R. Mohler III · Christopher J. O'Donnell · Walter Palmas · Weihong Tang · Russell P. Tracy · Alexander P. Reiner · Guillaume Lettre

Received: 28 September 2010 / Accepted: 25 November 2010 / Published online: 12 December 2010

Abstract Red blood cell, white blood cell, and platelet measures, including their count, sub-type and volume, are important diagnostic and prognostic clinical parameters for several human diseases. To identify novel loci associated with hematological traits, and compare the architecture of these phenotypes between ethnic groups, the CARe Project genotyped 49,094 single nucleotide polymorphisms (SNPs) that capture variation in ~2,100 candidate genes in DNA of 23,439 Caucasians and 7,112 African Americans from five population-based cohorts. We found strong novel associations between erythrocyte phenotypes and the glucose-6 phosphate dehydrogenase (*G6PD*) A-allele in African Americans (rs1050828, $P < 2.0 \times 10^{-13}$, T-allele associated with lower red blood cell count, hemoglobin, and hematocrit, and higher mean corpuscular volume), and between platelet count and a SNP at the tropomyosin-4 (*TPM4*) locus (rs8109288, $P = 3.0 \times 10^{-7}$ in Caucasians; $P = 3.0 \times 10^{-7}$ in African Americans, T-allele associated

A. P. Reiner and G. Lettre co-directed the study.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-010-0925-1) contains supplementary material, which is available to authorized users.

K. S. Lo · G. Galarneau · G. Lettre (✉)
Montreal Heart Institute, 5000 Bélanger Street,
Montreal, QC H1T 1C8, Canada
e-mail: guillaume.lettre@mhi-humangenetics.org

J. G. Wilson
Department of Medicine, G.V. (Sonny) Montgomery V.A. Medical Center and the University of Mississipi Medical Center,
Jackson, MS 39216, USA

L. A. Lange
Department of Genetics, University of North Carolina,
5112 Genetic Medicine Building, Chapel Hill,
NC 27599-7264, USA

A. R. Folsom · W. Tang
Division of Epidemiology and Community Health,
University of Minnesota, Minneapolis, MN 55454, USA

S. K. Ganesh
Division of Cardiovascular Medicine,
Department of Internal Medicine, The University of Michigan,
Ann Arbor, MI 48109, USA

S. F. A. Grant · B. J. Keating
Center for Applied Genomics, Division of Human Genetics,
Children's Hospital of Philadelphia Research Institute,
Philadelphia, PA 19104, USA

S. F. A. Grant
Department of Pediatrics, University of Pennsylvania
School of Medicine, Philadelphia, PA 19104, USA

S. A. McCarroll
Department of Genetics, Harvard Medical School,
Boston, MA 02115, USA

S. A. McCarroll
Broad Institute, Seven Cambridge Center,
Cambridge, MA 02142, USA

E. R. Mohler III
Cardiovascular Division, Vascular Medicine Section,
Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

C. J. O'Donnell
Division of Intramural Research, National Heart,
Lung, and Blood Institute (NHLBI), Bethesda, MD 20892, USA

C. J. O'Donnell
NHLBI's Framingham Heart Study, Framingham,
MA 01702, USA

C. J. O'Donnell
Cardiology Division, Massachusetts General Hospital,
Harvard Medical School, Boston, MA 02114, USA

with lower platelet count). We strongly replicated many genetic associations to blood cell phenotypes previously established in Caucasians. A common variant of the α-*globin* (*HBA2-HBA1*) locus was associated with red blood cell traits in African Americans, but not in Caucasians (rs1211375, $P < 7 \times 10^{-8}$, A-allele associated with lower hemoglobin, mean corpuscular hemoglobin, and mean corpuscular volume). Our results show similarities but also differences in the genetic regulation of hematological traits in European- and African-derived populations, and highlight the role of natural selection in shaping these differences.

Introduction

Blood cell counts are important clinical parameters: they are altered in many human diseases (e.g., cancers, infections and inflammation), and strongly modulate severity in primary blood disorders (e.g., the hemoglobinopathies). Genome-wide association studies (GWAS) in individuals of European ancestry have identified >30 loci that carry common DNA polymorphisms associated with blood cell numbers [including red blood cells (RBC), white blood cells (WBC), WBC sub-types, and platelets (PLT)] and related phenotypes [hematocrit (Hct), hemoglobin (Hb), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), and mean platelet volume (MPV)] (see Supplementary Table 1 for a definition of these different hematological traits and how they are derived) (Ganesh et al. 2009; Gudbjartsson et al. 2009; Meisinger et al. 2009; Soranzo et al. 2009a, b; Uda et al. 2008). Recent GWAS have also reported genetic associations between common SNPs and blood cell parameters in Japanese populations (Kamatani et al. 2010; Okada et al. 2010). Finally, genetic variation at the Duffy antigen receptor for chemokines (*DARC*) locus explains reduced total WBC and neutrophil levels in African Americans, compared to European Americans (Nalls et al. 2008; Reich et al. 2009).

To identify novel loci associated with blood cell counts and related indices, and to compare the genetic architecture of these quantitative traits in two different ethnic groups, we analyzed genetic association between each of 12 hematologic traits and ~175,000 SNPs that were genotyped or imputed in 23,439 Caucasians and 7,112 African Americans. These SNPs had been selected to capture variation across multiple ethnic groups (including African-derived populations) in ~2,100 candidate genes for heart, lung, and blood disorders (Keating et al. 2008). Here, we present evidence that the canonical glucose-6 phosphate dehydrogenase (*G6PD*) A-allele (rs1050828) implicated in *G6PD* deficiency (MIM #305900) and malaria resistance (Guindo et al. 2007; Ruwende et al. 1995; Tishkoff et al. 2001) also associates with RBC, Hct, Hb, and MCV variation ($P < 2 \times 10^{-13}$) in African Americans. We also identified a novel association between rs8109288 in the *TPM4* gene and PLT count in Caucasians ($P = 3.0 \times 10^{-7}$) and African Americans ($P = 3.0 \times 10^{-7}$). In Caucasians, we replicated the association between 13 distinct loci and hematological traits (Ganesh et al. 2009; Soranzo et al. 2009b). In African Americans, we found strong evidence of association between the *DARC* locus and WBC count (Nalls et al. 2008; Reich et al. 2009), and between the α-*globin* (*HBA2-HBA1*) locus and RBC count.

Materials and methods

Ethics statement

All participants gave informed written consent. The CARe project is approved by the ethics committees of the participating studies and of the Massachusetts Institute of Technology. This project was also reviewed and approved by the Montreal Heart Institute's ethics committee.

Samples and genotyping

Phenotypes from 23,439 Caucasians and 7,112 African Americans from the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in young Adults (CARDIA) study, the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), and the Jackson Heart Study (JHS) were analyzed in this study. A detailed description of each cohort can be found in the Supplementary Information.

All samples were genotyped by the Genetic Analysis Platform at the Broad Institute using the ITMAT-Broad-CARe (IBC) Illumina iSELECT array according to the

W. Palmas
Department of Medicine, Columbia University,
622 West 168th Street, Ph 9 East, 107,
New York, NY 10032, USA

R. P. Tracy
Departments of Pathology and Biochemistry,
University of Vermont, 208 S. Park Drive, suite 2,
Colchester, VT 05446, USA

A. P. Reiner (✉)
Department of Epidemiology, University of Washington,
Box 357236, Seattle, WA 98195, USA
e-mail: apreiner@u.washington.edu

G. Lettre
Département de Médecine, Université de Montreal, C.P. 6128,
succursale Centre-ville, Montreal, QC H3C 3J7, Canada

manufacturer's recommendations (Keating et al. 2008). Genotypes were called using Beadstudio (Illumina) and the calling cluster Id CVDSNP55v1_A.EGT. Quality control filters applied are summarized in Supplementary Tables 2–3.

Genotype imputation

Imputation was performed using MACH 1.0.16 (ref. Li et al. 2009). MACH requires phased reference haplotypes to perform imputation. For European Americans, we used the reference haplotypes from the Northern European CEU population from HapMap phase 2. For the African Americans, a combined CEU + YRI reference panel was created using HapMap phase 2 data (Kang et al. 2010). This panel includes SNPs segregating in both CEU and YRI, as well as SNPs segregating in one panel and monomorphic and nonmissing in the other. For both European Americans and African Americans, imputation was performed in two steps. For the first step, 300 individuals were randomly extracted to generate recombination and error rate estimates. In the second step, these rates were used to impute all individuals across the entire reference panel. Imputation results were filtered at an rsq_hat threshold $\geq$0.6 and a MAF threshold $\geq$1%.

Association testing

Methods used to measure the blood traits analyzed have been described previously for ARIC (Ganesh et al. 2009), CARDIA (Shimakawa and Bild 1993), CHS (Ganesh et al. 2009), FHS (Ganesh et al. 2009), and JHS (Reich et al. 2009). Because we use linear regression to analyze genotype–phenotype associations (see below), the distribution of the quantitative traits analyzed needs to be normal. Trait values were normalized into *Z*-scores using inverse normal transformation after accounting for gender, age, age-squared, and recruitment center (when available). Inverse normal transformation uses ranks to fit all phenotypic residuals into a perfectly normal distribution such that even individuals with extreme phenotype values ("outliers") can be kept in the analysis. We excluded individuals with blood cancers or known pregnancy at the time of visit.

For all cohorts but FHS, analysis was performed in PLINK (Purcell et al. 2007) using linear regression under an additive genetic model. For FHS, we modeled the family structure in the association tests using a linear mixed effects (LME) model implemented in R (Chen and Yang 2010). We tested an additive genetic model and included as covariates the first ten principal components. For imputed SNPs, dosage information (bound between 0.0 and 2.0) was used as predictor.

Association results were combined within ethnic group using the inverse variance method, as implemented in the software METAL (Willer et al. 2010). Individual study results were corrected using genomic control; meta-analytic results were also scaled using genomic control (Devlin and Roeder 1999).

Results

The National Heart, Lung and Blood Institute (NHLBI)-funded Candidate gene Association Resource (CARe) Project genotyped >40,000 participants from nine population-based cohorts to identify genetic associations with cardiovascular, pulmonary, hematologic, and sleep-related traits (Musunuru et al. 2010). In the CARe dataset, blood indices were available for up to 23,439 Caucasians and 7,112 African Americans from the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in young Adults (CARDIA) study, the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), and the Jackson Heart Study (JHS) (Table 1). All CARe samples were genotyped on the ITMAT-Broad-CARe (IBC) platform, which interrogates 49,094 SNPs (including many rare non-synonymous SNPs) to capture genetic variation in $\sim$2,100 candidate genes in multiple ethnic groups (Keating et al. 2008). Genotype data were processed using stringent quality-control filters (Supplementary Tables 2–3), and we used genotype imputation to increase coverage ("Materials and methods"). Phenotypes were analyzed as quantitative traits under an additive genetic model using a linear regression framework (Purcell et al. 2007). For each analysis, we also included as covariates the first ten principal components to correct for global admixture in African Americans and possible population stratification. Within each ethnic group, results were combined by meta-analysis (Willer et al. 2010). Individual study results, as well as meta-analytic results, were scaled using genomic control (Devlin and Roeder 1999).

Meta-analysis results for the 12 phenotypes analyzed in Caucasians and African Americans are summarized in quantile–quantile (QQ) plots (Supplementary Figs. 1–4). Except for WBC count in African Americans, the inflation factors observed are near unity (range λ_{GC} 0.948–1.065) (Supplementary Table 4), suggesting that our results are not markedly inflated by confounding factors. However, in African Americans, we observed a high inflation factor for WBC count ($\lambda_{GC} = 1.113$). Most of the inflation is due to the *DARC* locus on chromosome 1, which was originally identified as a bona fide WBC count locus by admixture mapping (Nalls et al. 2008). When we exclude SNPs on chromosome 1, the inflation factor for WBC count in

Table 1 Description of the hematological traits in the CARe cohorts

Phenotypes	ARIC Caucasians	ARIC African Americans	CARDIA Caucasians	CARDIA African Americans	CHS Caucasians	CHS African Americans	FHS (Caucasians)	JHS (African Americans)
Age	54.4 ± 5.7	53.6 ± 5.9	25.5 ± 3.4	24.4 ± 3.8	72.8 ± 5.6	72.9 ± 5.6	44.2 ± 10.1	54.2 ± 12.9
Male/female	4,456/5,132	1,118/1,911	672/771	526/769	1,732/2,220	280/472	3,380/4,108	803/1,231
White blood cell ($\times 10^9$/l)	6.30 ± 2.00 (9,493)	5.67 ± 1.97 (2,896)	6.21 ± 1.75 (1,430)	5.91 ± 2.00 (1,275)	6.35 ± 2.00 (3,842)	5.88 ± 2.57 (713)	6.45 ± 1.79 (2,445)	5.64 ± 2.31 (1,881)
Basophils (%)	0.53 ± 0.66 (6,249)	0.70 ± 0.71 (2,843)	0.68 ± 0.71 (1,143)	0.81 ± 0.71 (870)	NA	NA	NA	0.59 ± 0.36 (1,766)
Eosinophils (%)	2.20 ± 2.11 (6,250)	2.94 ± 2.49 (2,881)	2.78 ± 2.20 (1,335)	2.79 ± 2.14 (1,137)	NA	NA	NA	2.60 ± 2.25 (1,852)
Lymphocytes (%)	30.77 ± 7.82 (6,738)	39.56 ± 11.53 (2,890)	35.19 ± 9.07 (1,430)	39.26 ± 11.02 (1,275)	NA	NA	NA	35.42 ± 9.67 (1,877)
Monocytes (%)	5.90 ± 2.46 (6,698)	6.24 ± 3.06 (2,888)	5.20 ± 2.59 (1,423)	5.40 ± 2.60 (1,266)	NA	NA	NA	7.20 ± 2.25 (1,875)
Neutrophils (%)	60.76 ± 8.33 (6,706)	47.48 ± 12.71 (2,890)	54.48 ± 9.82 (1,050)	51.66 ± 11.69 (753)	NA	NA	NA	53.95 ± 10.23 (1,877)
Red blood cell ($\times 10^{12}$/l)	NA	NA	4.82 ± 0.50 (1,430)	4.77 ± 0.57 (1,275)	NA	NA	4.79 ± 0.44 (2,445)	4.53 ± 0.51 (1,881)
Hemoglobin (g/dl)	14.13 ± 1.32 (9,492)	13.18 ± 1.51 (2,896)	14.58 ± 1.36 (1,430)	13.82 ± 1.70 (1,275)	14.13 ± 1.33 (3,842)	13.41 ± 1.75 (713)	14.61 ± 1.37 (2,445)	13.04 ± 1.48 (2,004)
Hematocrit (%)	42.19 ± 3.75 (9,493)	40.22 ± 4.46 (2,896)	42.97 ± 4.13 (1,430)	41.20 ± 4.53 (1,275)	42.21 ± 3.87 (3,842)	39.87 ± 4.01 (713)	44.08 ± 3.83 (2,445)	39.33 ± 4.18 (2,004)
Mean corpuscular hemoglobin (pg)	NA	NA	NA	NA	NA	NA	30.53 ± 1.76 (2,445)	28.90 ± 2.52 (1,881)
Mean corpuscular volume (fl)	90.50 ± 4.26 (9,194)	86.67 ± 6.26 (2,769)	NA	NA	NA	NA	92.00 ± 5.01 (2,445)	87.02 ± 6.42 (1,881)
Platelet ($\times 10^9$/l)	258.22 ± 75.82 (9,465)	256.54 ± 67.12 (2,895)	263.69 ± 61.16 (1,429)	280.15 ± 72.03 (1,274)	250.21 ± 76.30 (3,820)	245.51 ± 71.80 (707)	NA	251.47 ± 64.94 (2,005)

Mean ± standard deviation is given for each phenotype. The numbers in parenthesis correspond to the number of samples with available phenotype

NA phenotype not measured or not available because it was not submitted to the CARe data coordinating center

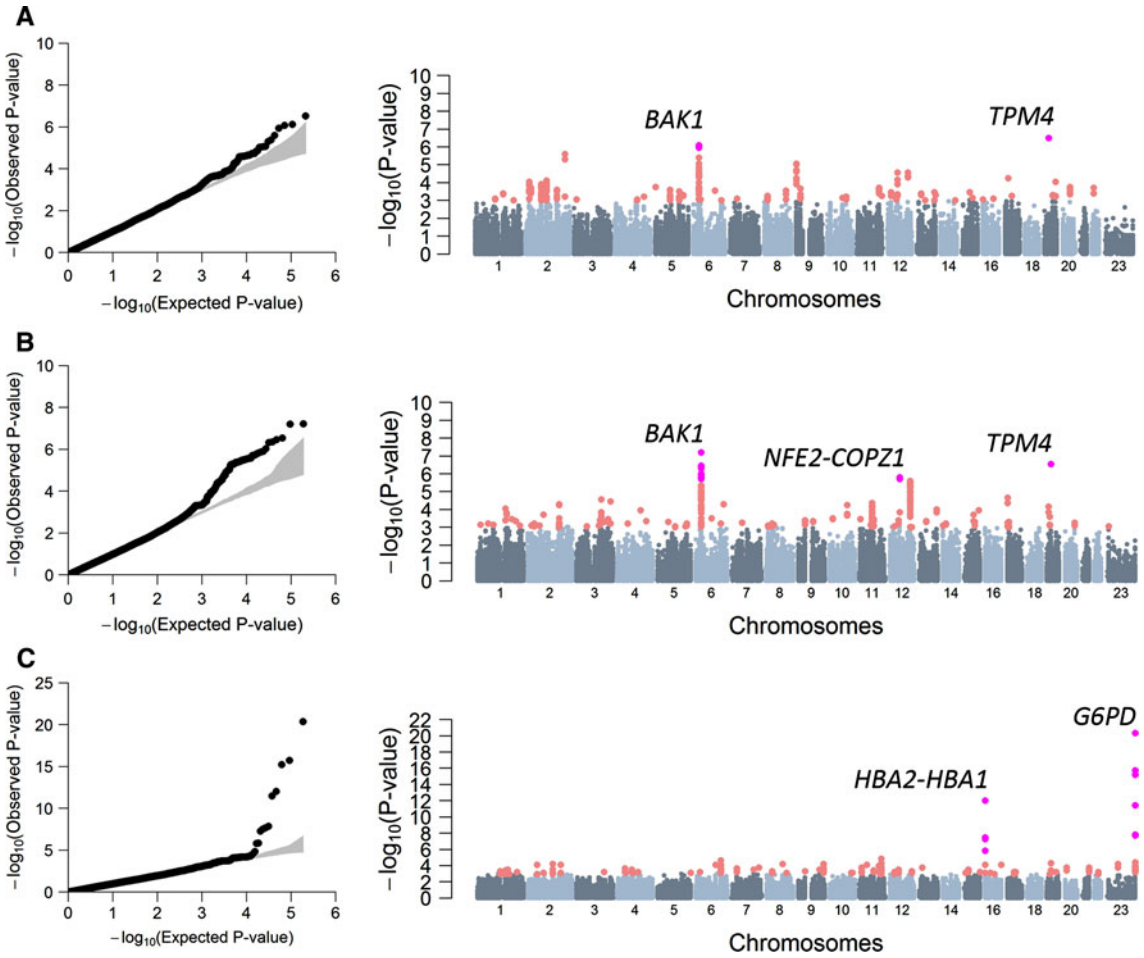


Fig. 1 Association results for selected hematological taits. Quantile–quantile (QQ) plots (*left panels*) and Manhattan plots (*right panels*) of genetic association results for platelet counts in Caucasians (**a**) and African Americans (**b**), and red blood cell counts in African Americans (**c**)

African Americans is reduced to $\lambda_{GC} = 1.054$ (Supplementary Fig. 5).

In this study, we consider a threshold of $P \leq 2 \times 10^{-6}$ as significant after accounting for the number of independent loci tested on the IBC platform (see Supplementary Information for a discussion of this statistical threshold). Our analysis highlighted novel genetic associations to hematological traits at $P \leq 2 \times 10^{-6}$, including two loci that reach the generally accepted threshold to declare genome-wide significance ($P \leq 5 \times 10^{-8}$) (Fig. 1; Table 2). In African Americans, we found that the missense SNP rs1050828 (Val68Met) in the *G6PD* gene is associated with RBC count, Hb, Hct, and MCV (all with *P* values $<2.0 \times 10^{-13}$). This SNP corresponds to the canonical *G6PD* A-allele, shown to cause partial *G6PD* deficiency (MIM #305900) and resistance to malaria (Guindo et al. 2007; Ruwende et al. 1995; Tishkoff et al. 2001). The *G6PD* A-allele is associated with decreased RBC count, Hb, and Hct, and increased MCV. As expected, because *G6PD* is X-linked, effect sizes for rs1050828 on RBC trait variation were stronger in males.

The second novel genome-wide significant association that we identified is between an intronic SNP in *TPM4* and PLT count. The A-allele at rs8109288 was associated with a decreased PLT count in both Caucasians ($P = 3.0 \times 10^{-7}$) and African Americans ($P = 3.0 \times 10^{-7}$) (Fig. 1; Table 2). The FHS does not have a standard PLT count available, but PLT count has been measured in platelet-rich plasma (PRP) as part of a study assessing in vitro platelet aggregation responses (O'Donnell et al. 2001). The A-allele at rs8109288 was associated with a decreased PLT count in PRP in the FHS ($P = 0.005$), providing additional, independent evidence that this sequence variant at the *TPM4* locus is linked to a variant that affects PLT number. *TPM4* is one of the four human tropomyosin genes, whose protein products play a role in cytoskeletal functions. An intronic SNP in *TPM1* (rs11071720) was recently shown to associate with MPV in Caucasians (Soranzo et al. 2009b), and a proxy of this SNP, rs3803499 [$r^2 = 0.25$ in HapMap Northern European (CEU) samples], was genotyped on the IBC array and is nominally

Table 2 Novel loci identified in this study that are associated with hematological traits in Caucasians and/or African Americans ($P \leq 2 \times 10^{-6}$)

Trait	Chr. (Pos.)	SNP	Locus	Annotation	Effect allele (effect allele frequency)	BETA (SE)	*P* value	Heterogeneity I^2 (%)
Caucasians								
Mean corpuscular volume	19 (61383413)	rs7258661	*GALP*	Intron	A (0.14)	0.097 (0.019)	8.3×10^{-7}	0
Platelets	19 (16046559)	rs8109288	*TPM4*	Intron	A (0.02)	−0.232 (0.044)	3.0×10^{-7}	0
African Americans								
Monocytes	2 (182032011)	rs1375493	*ITGA4*	Intron	A (0.57)	0.097 (0.019)	5.1×10^{-7}	0
Red blood cells	23 (153417411)	rs1050828	*G6PD*	Missense	T (0.11)	−0.415 (0.044)	4.3×10^{-21}	0
Hemoglobin	16 (3041640)	rs7199221	*MMP25*	Intron	A (0.44)	−0.084 (0.017)	1.3×10^{-6}	0
	23 (153417411)	rs1050828	*G6PD*	Missense	T (0.11)	−0.241 (0.030)	1.2×10^{-15}	0
Hematocrit	23 (153417411)	rs1050828	*G6PD*	Missense	T (0.11)	−0.222 (0.030)	1.2×10^{-13}	0
Mean corpuscular volume	11 (4970042)	rs16908114	*MMP26*	Missense	A (0.94)	0.227 (0.044)	4.9×10^{-7}	0
	16 (5634379)	rs12933048	16p13	Intergenic	A (0.71)	−0.126 (0.024)	5.0×10^{-7}	0
	23 (153417411)	rs1050828	*G6PD*	Missense	T (0.11)	0.340 (0.038)	3.0×10^{-18}	0
Platelets	12 (52998575)	rs10876550	*NFE2-COPZ1*	Intergenic	A (0.88)	0.135 (0.027)	1.6×10^{-6}	0
	19 (16046559)	rs8109288	*TPM4*	Intron	A (0.09)	−0.159 (0.030)	3.0×10^{-7}	0

Genomic positions and annotations are given using NCBI build 36.1. Effect allele is always on the forward strand. Effect size (BETA) and standard error (SE) are given in *Z*-score units after inverse normal transformation. *P* values are scaled using genomic control

associated with PLT count in the CARe cohorts ($P = 0.013$ in Caucasians, $P = 0.007$ in African Americans). In the JHS, the only CARe cohort with MPV available, the association between *TPM4* rs8109288 and MPV was strong ($P = 9.7 \times 10^{-5}$; the A-allele is associated with increased MPV). Nineteen SNPs in the two remaining tropomyosin genes, *TPM2* and *TPM3*, were also genotyped on the IBC array. None of these SNPs, or nearby imputed SNPs, were convincingly associated with PLT phenotypes in Caucasians or African Americans. Therefore, at least two tropomyosin genes harbor common genetic polymorphisms associated with PLT count and volume.

Our analysis strategy was validated by the replication of several associations to blood cell phenotypes reported previously. We replicated associations of 14 loci with one or more hematological traits (Table 3). Obviously, our replication is limited to the known loci covered by the IBC platform: loci that were interrogated by the IBC array and did not replicate, or loci not genotyped on this platform are listed in Supplementary Tables 5 and 6, respectively. Furthemore, because Caucasian samples from the ARIC, CHS, and FHS cohorts overlap with samples used in the CHARGE meta-analysis (Ganesh et al. 2009), replication of the CHARGE findings in the CARe meta-analysis does not provide independent confirmatory evidences. In some cases, we identified the same SNPs as previous studies, such as the missense SNP rs1800562 in *HFE* associated with Hb, Hct, and MCV in Caucasians, or rs2814778 located in the 5′ untranslated region (UTR) of *DARC*, strongly associated with WBC and neutrophil levels in African Americans (Table 3). Of particular interest, the missense SNP rs3184504 in *SH2B3* (Trp262Arg) illustrates an extreme case of pleiotropy: this SNP has now been associated with Hb and Hct (Table 3) (Ganesh et al. 2009), PLT count (Kamatani et al. 2010; Soranzo et al. 2009b), eosinophil count and myocardial infarction (Gudbjartsson et al. 2009), blood pressure and hypertension (Levy et al. 2009), celiac disease (Hunt et al. 2008), and type 1 diabetes (Todd et al. 2007) in Caucasians. *SH2B3* encodes Lnk, a negative regulator of hematopoiesis, and it is possible that genetic variation in this gene affects common disease risk indirectly by modulating properties of the three major blood cell types.

Discussion

We have used a genotyping array that covers common genetic variation in ~2,100 candidate genes to identify loci associated with blood cell counts and related indices in Caucasians and African Americans. SNPs on this array were selected using a "cosmopolitan" tagging approach

Table 3 Replication of loci previously shown to carry SNPs associated with hematological traits in Caucasians and/or African Americans

Trait	Chr. (Pos.)	SNP	Locus	Annotation	Effect allele (effect allele frequency)	BETA (SE)	*P* value	Heterogeneity I^2 (%)	Reference
Caucasians									
White blood cells	6 (32558067)	rs5020946	6p21	Intergenic	T (0.42)	0.064 (0.011)	1.9×10^{-8}	0	Kamatani et al. (2010)
	6 (135559689)	rs12660713	*MYB*	Intron	A (0.11)	−0.062 (0.017)	0.00044	0	Kamatani et al. (2010)
	7 (92246306)	rs445	*CDK6*	Intron	T (0.12)	−0.061 (0.017)	0.00041	0	Kamatani et al. (2010)
	17 (35462466)	rs8065443	*CSF3*	Intron	A (0.42)	0.124 (0.014)	1.8×10^{-19}	0	Kamatani et al. (2010), Soranzo et al. 2009b)
Lymphocytes	17 (35462466)	rs8065443	*CSF3*	Intron	A (0.42)	−0.114 (0.020)	1.4×10^{-8}	0	Kamatani et al. (2010)
Neutrophils	17 (35462466)	rs8065443	*CSF3*	Intron	A (0.42)	0.119 (0.020)	3.8×10^{-9}	0	Okada et al. (2010)
Red blood cells	6 (135545478)	rs210962	*MYB*	Intron	T (0.24)	0.105 (0.029)	0.00047	0	Ganesh et al. (2009), Kamatani et al. (2010)
	7 (100062967)	rs2075674	*TFR2*	Synonymous	A (0.21)	0.132 (0.030)	1.6×10^{-5}	0	Soranzo et al. (2009b)
	7 (100159464)	rs551238	*EPO*	Near gene 3′	T (0.60)	0.069 (0.014)	1.3×10^{-6}	0	Ganesh et al. (2009)
Hemoglobin	6 (26201120)	rs1800562	*HFE*	Missense	A (0.06)	0.187 (0.023)	1.4×10^{-15}	19.4	Ganesh et al. (2009), Soranzo et al. (2009b)
	7 (151045974)	rs10224002	*PRKAG2*	Intron	A (0.71)	0.079 (0.012)	1.6×10^{-10}	0	Ganesh et al. (2009)
	9 (135143696)	rs651007	*ABO*	Intergenic	T (0.22)	−0.045 (0.013)	0.00096	0	Kamatani et al. (2010)
	12 (110368991)	rs3184504	*SH2B3*	Missense	T (0.50)	0.057 (0.011)	3.2×10^{-7}	0	Soranzo et al. (2009b)
Hematocrit	6 (26201120)	rs1800562	*HFE*	Missense	A (0.06)	0.146 (0.023)	2.5×10^{-10}	0	Ganesh et al. (2009)
	6 (135577714)	rs6936293	*MYB*	Intron	T (0.12)	−0.059 (0.017)	0.00055	0	Ganesh et al. (2009), Kamatani et al. (2010)
	7 (100051951)	rs7786877	*TFR2*	Intergenic	A (0.75)	−0.069 (0.013)	4.3×10^{-7}	0	Ganesh et al. (2009)
	7 (151044127)	rs10224210	*PRKAG2*	Intron	T (0.71)	0.071 (0.012)	5.0×10^{-9}	0	Ganesh et al. (2009)
	12 (110368991)	rs3184504	*SH2B3*	Missense	T (0.50)	0.053 (0.011)	1.5×10^{-6}	0	Soranzo et al. (2009b)
Mean corpuscular volume	3 (197265106)	rs557527	*TFRC*	Intron	A (0.33)	−0.050 (0.014)	0.00064	0	Ganesh et al. (2009), Kamatani et al. (2010)
	6 (26201120)	rs1800562	*HFE*	Missense	A (0.06)	0.233 (0.028)	6.8×10^{-16}	0	Ganesh et al. (2009), Soranzo et al. (2009b)
	6 (135526787)	rs17706858	*MYB*	Intergenic	T (0.14)	0.134 (0.022)	3.9×10^{-9}	0	Ganesh et al. (2009), Kamatani et al. (2010), Soranzo et al. (2009b))
	7 (100073906)	rs7385804	*TFR2*	Intron	A (0.62)	0.076 (0.014)	1.6×10^{-7}	0	Ganesh et al. (2009)
	10 (45259824)	rs2288619	*ALOX5*	Intron	T (0.07)	0.100 (0.028)	0.00041	23.3	Kamatani et al. (2010)
	22 (49308720)	rs2782	*ECGF1*	utr 3′	T (0.39)	−0.070 (0.014)	1.0×10^{-6}	74.5	Ganesh et al. (2009)
Platelets	6 (33654476)	rs5745582	*BAK1*	Intron	T (0.21)	0.076 (0.015)	7.6×10^{-7}	0	Kamatani et al. (2010), Soranzo et al. (2009b)

Table 3 continued

Trait	Chr. (Pos.)	SNP	Locus	Annotation	Effect allele (effect allele frequency)	BETA (SE)	*P* value	Heterogeneity I^2 (%)	Reference
African Americans									
White blood cells	1 (157441307)	rs2814778	*DARC*	utr 5′	T (0.19)	0.811 (0.036)	1.4×10^{-103}	0	Nalls et al. (2008), Reich et al. (2009)
Basophils	1 (157441307)	rs2814778	*DARC*	utr 5′	T (0.18)	−0.334 (0.044)	1.1×10^{-13}	69.3	Nalls et al. (2008), Reich et al. (2009)
Lymphocytes	1 (157441307)	rs2814778	*DARC*	utr 5′	T (0.18)	−0.786 (0.041)	4.1×10^{-82}	0	Nalls et al. (2008), Reich et al. (2009)
	7 (92246306)	rs445	*CDK6*	Intron	T (0.20)	0.130 (0.023)	2.0×10^{-8}	0	Kamatani et al. (2010)
Monocytes	1 (160296531)	rs4657139	*DARC*	Intergenic	A (0.87)	0.183 (0.030)	3.6×10^{-9}	0	Nalls et al. (2008), Reich et al. (2009)
Neutrophils	1 (157441307)	rs2814778	*DARC*	utr 5′	T (0.18)	0.856 (0.043)	5.5×10^{-85}	0	Nalls et al. (2008), Reich et al. (2009)
Hemoglobin	16 (180281)	rs1211375	*HBA2-HBA1*	Intron of *LUC7L*	A (0.28)	−0.103 (0.019)	6.1×10^{-8}	0	Kamatani et al. (2010)
Mean corpuscular hemoglobin	16 (180281)	rs1211375	*HBA2-HBA1*	Intron of *LUC7L*	A (0.27)	−0.283 (0.040)	7.4×10^{-13}	0	Kamatani et al. (2010)
Mean corpuscular volume	11 (5259609)	rs2213170	*HBB*	Intergenic	A (0.13)	−0.115 (0.032)	0.00029	0	Galanello et al. (1979)
	16 (180281)	rs1211375	*HBA2-HBA1*	Intron of *LUC7L*	A (0.28)	−0.293 (0.024)	1.4×10^{-33}	0	Kamatani et al. (2010)
Platelet	6 (33604692)	rs449242	*BAK1*	Intergenic	T (0.33)	−0.126 (0.023)	6.2×10^{-8}	0	Kamatani et al. (2010), Soranzo et al. (2009b)

To be included in this table, we required a $P \leq 0.001$. Genomic positions and annotations are given using NCBI build 36.1. Effect allele is always on the forward strand. Effect size (BETA) and standard error (SE) are given in *Z*-score units after normal inverse transformation. *P* values are scaled using genomic control

such that common variants in these candidate genes should be covered similarly in these two ethnic groups (Keating et al. 2008). We analyzed 12 phenotypes and identified two new loci that reach genome-wide significance: *G6PD* rs1050828 is associated with RBC count, Hb, Hct, and MCV in African Americans, and *TPM4* rs8109288 is associated with PLT count in Caucasians and African Americans (Table 2). Since clinical processes such as iron deficiency or inflammation can influence hematological traits, we sought to confirm that the associations observed at *G6PD* and *TPM4* were independent of these conditions. When we adjusted our analyses for iron levels [using ferritin or iron levels, or total-iron binding capacity (TIBC)] or inflammation [using C-reactive protein (CRP) levels], association results remained largely unchanged (data not shown), suggesting that the associations at *G6PD* and *TPM4* are independent of iron and inflammation status. In addition to the novel associations at *G6PD* and *TPM4*, we replicated 36 previously reported associations (Table 3).

Glucose-6-phosphate dehydrogenase protects red blood cells against oxidative damage. Inherited deficiency of glucose-6-phosphate dehydrogenase is an X-linked enzymopathy that has a higher prevalence in areas of the world where malaria is endemic. Many variants of *G6PD* have been described with wide ranging levels of enzyme activity and associated clinical symptoms. Rare severe mutations in *G6PD* have been linked to neonatal jaundice and to acute and chronic hemolytic anemia (including congenital nonspherocytic hemolytic anemia) in the presence of oxidative stress (Cappellini and Fiorelli 2008). To our knowledge, our study is the first to report associations between the mild *G6PD* A-variant and erythrocyte phenotypes in normal populations, although it was recently associated with Hb levels in sickle cell anemia patients (Nouraie et al. 2010).

We have conducted one of the first well-powered genetic association studies where several complex human traits are analyzed in two ethnic groups, allowing a direct comparison of the architecture of these phenotypes in individuals of European and African descent. The overlap in loci that control hematological traits in Caucasians and African Americans was small, with only SNPs at the *BAK1* and *TPM4* genes found consistently associated with the same trait at $P \leq 2 \times 10^{-6}$ (PLT count in both cases) (Soranzo et al. 2009b). This lack of overlap might be partially explained by the difference in sample sizes: 7,112 African Americans and 23,439 Caucasians were available in our analyses. Consequently, when more African American datasets become available, more loci will likely be shown to control blood cell phenotypes in both of these two ethnic groups. It is also possible that difference in allele frequencies might affect discovery power between Caucasians and African Americans.

However, it is also apparent that differing selective pressures have shaped the genetic regulation of hematological traits in European- and African-derived populations (see Supplementary Table 7 for iHS values). The *HFE* rs1800562 missense SNP (C282Y) is associated with erythrocyte phenotypes (Hb, Hct, MCV) (Ganesh et al. 2009; Soranzo et al. 2009b) and iron status (Benyamin et al. 2009), and causes hereditary hemochromatosis (MIM #235200) in Caucasians (Table 3). The minor A-allele of rs1800562 is absent in populations of African ancestry but relatively frequent in Caucasians (5–10%), and is located on a long haplotype of low diversity indicative of positive selection (Ajioka et al. 1997; Thomas et al. 1998). Similarly, it was recently shown that the *SH2B3* locus associated with variation in all three main blood cell type (RBC, WBC, and PLT) indices is also under natural selection (Soranzo et al. 2009b). Genetic variation at the *SH2B3* locus is also associated with activation of the innate immune system, suggesting a possible role in protection against bacterial infection (Zhernakova et al. 2010). In African Americans, the three major loci associated with WBC and RBC trait variations are *DARC*, *HBA2-HBA1*, and *G6PD*. These three loci are known to carry alleles that confer resistance to malaria and to be under strong positive selection in African-derived populations. The minor alleles for *DARC* rs2814778 and *G6PD* rs1050828 are common in African Americans (>10%) but extremely rare (or absent) in Caucasians. The minor A-allele for *HBA2-HBA1* rs1211375 is common in both Caucasians (35%) and African Americans (27%), but only associates with erythrocyte phenotypes in African Americans. A recent survey of structural variants in the human genome has shown a copy number polymorphism (CNVR6569) at the *α-globin* locus in HapMap samples from Yoruba, Nigeria (YRI) that is absent in HapMap CEU individuals (Conrad et al. 2009). Of the SNPs at the *α-globin* locus surveyed by the IBC array, rs1211375 is the best tag for CNVR6569 ($r^2 = 0.37$). Thus, it is likely that rs1211375 captures variation in the number of *α-globin* genes in African Americans. This is consistent with the clinical epidemiology of α-thalassemia, which is known to modulate RBC phenotypes. As cross-ethnic association studies are performed for additional phenotypes, it will be interesting to compare the genetic architecture of complex human diseases and traits between ethnic groups, and to assess the role of natural selection in shaping the differences.

Acknowledgments The authors wish to acknowledge the support of the National Heart, Lung, and Blood Institute and the contributions of the research institutions, study investigators, field staff and study participants in creating this resource for biomedical research. The grants and contracts that have supported CARe are listed at http://public.nhlbi.nih.gov/GeneticsGenomics/home/care.aspx. Additional support for this work was provided by the Fondation de l'Institut de Cardiologie de Montréal (to G.L.), and by NIH R01 HL71862-06 and ARRA N000949304 (to A.P.R.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest We declare no competing interests.

References

Ajioka RS, Jorde LB, Gruen JR, Yu P, Dimitrova D, Barrow J, Radisky E, Edwards CQ, Griffen LM, Kushner JP (1997) Haplotype analysis of hemochromatosis: evaluation of different linkage-disequilibrium approaches and evolution of disease chromosomes. Am J Hum Genet 60:1439–1447

Benyamin B, Ferreira MA, Willemsen G, Gordon S, Middelberg RP, McEvoy BP, Hottenga JJ, Henders AK, Campbell MJ, Wallace L, Frazer IH, Heath AC, de Geus EJ, Nyholt DR, Visscher PM, Penninx BW, Boomsma DI, Martin NG, Montgomery GW, Whitfield JB (2009) Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. Nat Genet 41:1173–1175

Cappellini MD, Fiorelli G (2008) Glucose-6-phosphate dehydrogenase deficiency. Lancet 371:64–74

Chen MH, Yang Q (2010) GWAF: an R package for genome-wide association analyses with family data. Bioinformatics 26:580–581

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME (2009) Origins and functional impact of copy number variation in the human genome. Nature 462:704–712

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Galanello R, Melis MA, Ruggeri R, Addis M, Scalas MT, Maccioni L, Furbetta M, Angius A, Tuveri T, Cao A (1979) Beta 0 thalassemia trait in Sardinia. Hemoglobin 3:33–46

Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A, Kuhnel B, Aspelund T, Yang Q, Tanaka T, Jaffe A, Bis JC, Verwoert GC, Teumer A, Fox CS, Guralnik JM, Ehret GB, Rice K, Felix JF, Rendon A, Eiriksdottir G, Levy D, Patel KV, Boerwinkle E, Rotter JI, Hofman A, Sambrook JG, Hernandez DG, Zheng G, Bandinelli S, Singleton AB, Coresh J, Lumley T, Uitterlinden AG, Vangils JM, Launer LJ, Cupples LA, Oostra BA, Zwaginga JJ, Ouwehand WH, Thein SL, Meisinger C, Deloukas P, Nauck M, Spector TD, Gieger C, Gudnason V, van Duijn CM, Psaty BM, Ferrucci L, Chakravarti A, Greinacher A, O'Donnell CJ, Witteman JC, Furth S, Cushman M, Harris TB, Lin JP (2009) Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet 41:1191–1198

Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, Thorleifsson G, Helgadottir H, Steinthorsdottir V, Stefansson H, Williams C, Hui J, Beilby J, Warrington NM, James A, Palmer LJ, Koppelman GH, Heinzmann A, Krueger M, Boezen HM, Wheatley A, Altmuller J, Shin HD, Uh ST, Cheong HS, Jonsdottir B, Gislason D, Park CS, Rasmussen LM, Porsbjerg C, Hansen JW, Backer V, Werge T, Janson C, Jonsson UB, Ng MC, Chan J, So WY, Ma R, Shah SH, Granger CB, Quyyumi AA, Levey AI, Vaccarino V, Reilly MP, Rader DJ, Williams MJ, van Rij AM, Jones GT, Trabetti E, Malerba G, Pignatti PF, Boner A, Pescollderungg L, Girelli D, Olivieri O, Martinelli N, Ludviksson BR, Ludviksdottir D, Eyjolfsson GI, Arnar D, Thorgeirsson G, Deichmann K, Thompson PJ, Wjst M, Hall IP, Postma DS, Gislason T, Gulcher J, Kong A, Jonsdottir I, Thorsteinsdottir U, Stefansson K (2009) Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. Nat Genet 41:342–347

Guindo A, Fairhurst RM, Doumbo OK, Wellems TE, Diallo DA (2007) X-linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. PLoS Med 4:e66

Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, Gwilliam R, Takeuchi F, McLaren WM, Holmes GK, Howdle PD, Walters JR, Sanders DS, Playford RJ, Trynka G, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, O'Morain C, Kennedy NP, Kelleher D, Pennington DJ, Strachan DP, McArdle WL, Mein CA, Wapenaar MC, Deloukas P, McGinnis R, McManus R, Wijmenga C, van Heel DA (2008) Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet 40:395–402

Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet 42:210–215

Kang SJ, Chiang CW, Palmer CD, Tayo BO, Lettre G, Butler JL, Hackett R, Adeyemo AA, Guiducci C, Berzins I, Nguyen TT, Feng T, Luke A, Shriner D, Ardlie K, Rotimi C, Wilks R, Forrester T, McKenzie CA, Lyon HN, Cooper RS, Zhu X, Hirschhorn JN (2010) Genome-wide association of anthropometric traits in African- and African-derived populations. Hum Mol Genet 19:2725–2738

Keating BJ, Tischfield S, Murray SS, Bhangale T, Price TS, Glessner JT, Galver L, Barrett JC, Grant SF, Farlow DN, Chandrupatla HR, Hansen M, Ajmal S, Papanicolaou GJ, Guo Y, Li M, Derohannessian S, de Bakker PI, Bailey SD, Montpetit A, Edmondson AC, Taylor K, Gai X, Wang SS, Fornage M, Shaikh T, Groop L, Boehnke M, Hall AS, Hattersley AT, Frackelton E, Patterson N, Chiang CW, Kim CE, Fabsitz RR, Ouwehand W, Price AL, Munroe P, Caulfield M, Drake T, Boerwinkle E, Reich D, Whitehead AS, Cappola TP, Samani NJ, Lusis AJ, Schadt E, Wilson JG, Koenig W, McCarthy MI, Kathiresan S, Gabriel SB, Hakonarson H, Anand SS, Reilly M, Engert JC, Nickerson DA, Rader DJ, Hirschhorn JN, Fitzgerald GA (2008) Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. PLoS ONE 3:e3583

Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, Aulchenko Y, Lumley T, Kottgen A, Vasan RS, Rivadeneira F, Eiriksdottir G, Guo X, Arking DE, Mitchell GF, Mattace-Raso FU, Smith AV, Taylor K, Scharpf RB, Hwang SJ, Sijbrands EJ, Bis J, Harris TB, Ganesh SK, O'Donnell CJ, Hofman A, Rotter JI, Coresh J, Benjamin EJ, Uitterlinden AG, Heiss G, Fox CS, Witteman JC, Boerwinkle E, Wang TJ, Gudnason V, Larson MG, Chakravarti A, Psaty BM, van Duijn CM (2009) Genome-wide association study of blood pressure and hypertension. Nat Genet 41:677–687

Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype imputation. Annu Rev Genomics Hum Genet 10:387–406

Meisinger C, Prokisch H, Gieger C, Soranzo N, Mehta D, Rosskopf D, Lichtner P, Klopp N, Stephens J, Watkins NA, Deloukas P, Greinacher A, Koenig W, Nauck M, Rimmbach C, Volzke H, Peters A, Illig T, Ouwehand WH, Meitinger T, Wichmann HE, Doring A (2009) A genome-wide association study identifies three loci associated with mean platelet volume. Am J Hum Genet 84:66–71

Musunuru K, Lettre G, Young T, Farlow DN, Pirrucello JP, Ejebe KG, Yang Q, Chen MH (2010) Candidate Gene Association Resource (CARe): design, methods, and proof of concept. Circ Cardiovasc Genet 3:267–275

Nalls MA, Wilson JG, Patterson NJ, Tandon A, Zmuda JM, Huntsman S, Garcia M, Hu D, Li R, Beamer BA, Patel KV, Akylbekova EL, Files JC, Hardy CL, Buxbaum SG, Taylor HA, Reich D, Harris TB, Ziv E (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

Nouraie M, Reading NS, Campbell A, Minniti CP, Rana SR, Luchtman-Jones L, Kato GJ, Gladwin MT, Castro OL, Prchal JT, Gordeuk VR (2010) Association of G6PD with lower haemoglobin concentration but not increased haemolysis in patients with sickle cell anaemia. Br J Haematol 150:218–225

O'Donnell CJ, Larson MG, Feng D, Sutherland PA, Lindpaintner K, Myers RH, D'Agostino RA, Levy D, Tofler GH (2001) Genetic and environmental contributions to platelet aggregation: the Framingham heart study. Circulation 103:3051–3056

Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, Daigo Y, Yamamoto K, Kubo M, Nakamura Y, Kamatani N (2010) Common variations in PSMD3-CSF3 and PLCB4 are associated with neutrophil count. Hum Mol Genet 19:2079–2085

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

Reich D, Nalls MA, Kao WH, Akylbekova EL, Tandon A, Patterson N, Mullikin J, Hsueh WC, Cheng CY, Coresh J, Boerwinkle E, Li M, Waliszewska A, Neubauer J, Li R, Leak TS, Ekunwe L, Files JC, Hardy CL, Zmuda JM, Taylor HA, Ziv E, Harris TB, Wilson JG (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

Ruwende C, Khoo SC, Snow RW, Yates SN, Kwiatkowski D, Gupta S, Warn P, Allsopp CE, Gilbert SC, Peschu N et al (1995) Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. Nature 376:246–249

Shimakawa T, Bild DE (1993) Relationship between hemoglobin and cardiovascular risk factors in young adults. J Clin Epidemiol 46:1257–1266

Soranzo N, Rendon A, Gieger C, Jones CI, Watkins NA, Menzel S, Doring A, Stephens J, Prokisch H, Erber W, Potter SC, Bray SL, Burns P, Jolley J, Falchi M, Kuhnel B, Erdmann J, Schunkert H, Samani NJ, Illig T, Garner SF, Rankin A, Meisinger C, Bradley JR, Thein SL, Goodall AH, Spector TD, Deloukas P, Ouwehand WH (2009a) A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function. Blood 113:3831–3837

Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, Willenborg C, Wright B, Chen L, Li M, Salo P, Voight BF, Burns P, Laskowski RA, Xue Y, Menzel S, Altshuler D, Bradley JR, Bumpstead S, Burnett MS, Devaney J, Doring A, Elosua R, Epstein SE, Erber W, Falchi M, Garner SF, Ghori MJ, Goodall AH, Gwilliam R, Hakonarson HH, Hall AS, Hammond N, Hengstenberg C, Illig T, Konig IR, Knouff CW, McPherson R, Melander O, Mooser V, Nauck M, Nieminen MS, O'Donnell CJ, Peltonen L, Potter SC, Prokisch H, Rader DJ, Rice CM, Roberts R, Salomaa V, Sambrook J, Schreiber S, Schunkert H, Schwartz SM, Serbanovic-Canic J, Sinisalo J, Siscovick DS, Stark K, Surakka I, Stephens J, Thompson JR, Volker U, Volzke H, Watkins NA, Wells GA, Wichmann HE, Van Heel DA, Tyler-Smith C, Thein SL, Kathiresan S, Perola M, Reilly MP, Stewart AF, Erdmann J, Samani NJ, Meisinger C, Greinacher A, Deloukas P, Ouwehand WH, Gieger C (2009b) A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet 41:1182–1190

Thomas W, Fullan A, Loeb DB, McClelland EE, Bacon BR, Wolff RK (1998) A haplotype and linkage disequilibrium analysis of the hereditary hemochromatosis gene region. Hum Genet 102:517–525

Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J, Piro A, Stoneking M, Tagarelli A, Tagarelli G, Touma EH, Williams SM, Clark AG (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293:455–462

Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, Wallace C, Howson JM, Guja C, Ionescu-Tirgoviste C, Simmonds MJ, Heward JM, Gough SC, Dunger DB, Wicker LS, Clayton DG (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 39:857–864

Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G, Piras MG, Sestu N, Lai S, Dei M, Mulas A, Crisponi L, Naitza S, Asunis I, Deiana M, Nagaraja R, Perseu L, Satta S, Cipollina MD, Sollaino C, Moi P, Hirschhorn JN, Orkin SH, Abecasis GR, Schlessinger D, Cao A (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

Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191

Zhernakova A, Elbers CC, Ferwerda B, Romanos J, Trynka G, Dubois PC, de Kovel CG, Franke L, Oosting M, Barisani D, Bardella MT, Joosten LA, Saavalainen P, van Heel DA, Catassi C, Netea MG, Wijmenga C (2010) Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. Am J Hum Genet 86:970–977