

# Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium

Santhi K Ganesh<sup>1–3,45\*</sup>, Neil A Zakai<sup>4,45</sup>, Frank J A van Rooij<sup>5,6,45</sup>, Nicole Soranzo<sup>7,8,45</sup>, Albert V Smith<sup>9,45</sup>, Michael A Nalls<sup>10,45</sup>, Ming-Huei Chen<sup>11,12</sup>, Anna Kottgen<sup>13</sup>, Nicole L Glazer<sup>14,15</sup>, Abbas Dehghan<sup>5,6</sup>, Brigitte Kuhnel<sup>16</sup>, Thor Aspelund<sup>9,17</sup>, Qiong Yang<sup>11,18</sup>, Toshiko Tanaka<sup>19,20</sup>, Andrew Jaffe<sup>13</sup>, Joshua C M Bis<sup>15</sup>, Germaine C Verwoert<sup>5,6,21</sup>, Alexander Teumer<sup>22</sup>, Caroline S Fox<sup>2,11</sup>, Jack M Guralnik<sup>23</sup>, Georg B Ehret<sup>3,24</sup>, Kenneth Rice<sup>25</sup>, Janine F Felix<sup>5,6</sup>, Augusto Rendon<sup>26</sup>, Gudny Eiriksdottir<sup>9</sup>, Daniel Levy<sup>2,11</sup>, Kushang V Patel<sup>23</sup>, Eric Boerwinkle<sup>27</sup>, Jerome I Rotter<sup>28</sup>, Albert Hofman<sup>5,6</sup>, Jennifer G Sambrook<sup>26</sup>, Dena G Hernandez<sup>29</sup>, Gang Zheng<sup>30</sup>, Stefania Bandinelli<sup>31</sup>, Andrew B Singleton<sup>10</sup>, Josef Coresh<sup>13</sup>, Thomas Lumley<sup>25</sup>, André G Uitterlinden<sup>5,6,21,32</sup>, Janine M vanGils<sup>33</sup>, Lenore J Launer<sup>23</sup>, L Adrienne Cupples<sup>11,8</sup>, Ben A Oostra<sup>5,6,34</sup>, Jaap-Jan Zwaginga<sup>35,36</sup>, Willem H Ouwehand<sup>26,7</sup>, Swee-Lay Thein<sup>37</sup>, Christa Meisinger<sup>16</sup>, Panos Deloukas<sup>7</sup>, Matthias Nauck<sup>38</sup>, Tim D Spector<sup>8</sup>, Christian Gieger<sup>16</sup>, Vilmundur Gudnason<sup>9,17</sup>, Cornelia M van Duijn<sup>5,6</sup>, Bruce M Psaty<sup>14,15,39–41</sup>, Luigi Ferrucci<sup>19</sup>, Aravinda Chakravarti<sup>3</sup>, Andreas Greinacher<sup>42,45</sup>, Christopher J O'Donnell<sup>2,11,43,45</sup>, Jacqueline C M Witteman<sup>5,6,45</sup>, Susan Furth<sup>44,45</sup>, Mary Cushman<sup>4,45</sup>, Tamara B Harris<sup>23,45</sup> & Jing-Ping Lin<sup>30,45</sup>

Measurements of erythrocytes within the blood are important clinical traits and can indicate various hematological disorders. We report here genome-wide association studies (GWAS) for six erythrocyte traits, including hemoglobin concentration (Hb), hematocrit (Hct), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC) and red blood cell count (RBC). We performed an initial GWAS in cohorts of the CHARGE Consortium totaling 24,167 individuals of European ancestry and replication in additional independent cohorts of the HaemGen Consortium totaling 9,456 individuals. We identified 23 loci significantly associated with these traits in a meta-analysis of the discovery and replication cohorts (combined *P* values ranging from  $5 \times 10^{-8}$  to  $7 \times 10^{-86}$ ). Our findings include loci previously associated with these traits (*HBS1L-MYB*, *HFE*, *TMPRSS6*, *TFR2*, *SPTA1*) as well as new associations (*EPO*, *TFRC*, *SH2B3* and 15 other loci). This study has identified new determinants of erythrocyte traits, offering insight into common variants underlying variation in erythrocyte measures.

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 nonhematologic diseases and mortality<sup>1–3</sup>.

Erythrocyte production and quality are under various environmental and genetic influences. Although environmental exposures, dietary intake of vitamins and iron, and the anemia associated with chronic diseases all contribute substantially to abnormalities of erythrocyte measures, the heritability of erythrocyte traits ranges from 40 to 90%<sup>4–6</sup>. Disorders of hemoglobin production and

hemoglobinopathies are some of the most common genetic diseases in the world. Some known low-frequency mendelian variants also lead to interindividual variability in erythrocyte traits among the general population<sup>7,8</sup>. Candidate gene studies have identified a few nonhemoglobin loci, including *EPOR* and *HBS1L*, related to variation in erythrocyte traits<sup>8–10</sup>. Early GWAS and linkage studies of erythrocyte measures, which have identified a few associations, such as that at chromosome 6q23, lacked statistical power for association and genetic resolution for testing competing hypotheses<sup>6,11–13</sup>.

To search for genetic determinants of erythrocyte traits in the general population, we carried out GWAS and meta-analysis within multiple community-based cohorts comprising the CHARGE Consortium<sup>14</sup>, followed by replication in additional independent cohorts from the HaemGen Consortium. 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 identified loci with blood pressure and hypertension.

\*A full list of author affiliations appears at the end of the paper.

Received 8 July; accepted 1 September; published online 11 October 2009; doi:10.1038/ng.466

**Table 1** CHARGE cohort description

	AGES	ARIC	CHS	FHS	RS	InCHIANTI
Number of individuals eligible for GWAS	3,219	8,127	3,275	3,381	5,523	1,206
Percent women	58	53	61	54	60	56
Mean age (years)	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 (pg)	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 (fl)	92.08 (4.49)	90.7 (4.22)	NA	90.7 (5.0)	88.29 (4.30)	90.22 (4.84)
RBC (1 M cell/mm <sup>3</sup> )	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	1989–1990	1971–1975	1990–1993	1998
Years of DNA collection	2002–2006	1987–1998	1989–1990	1996–1999	1990–1993	1998–2001

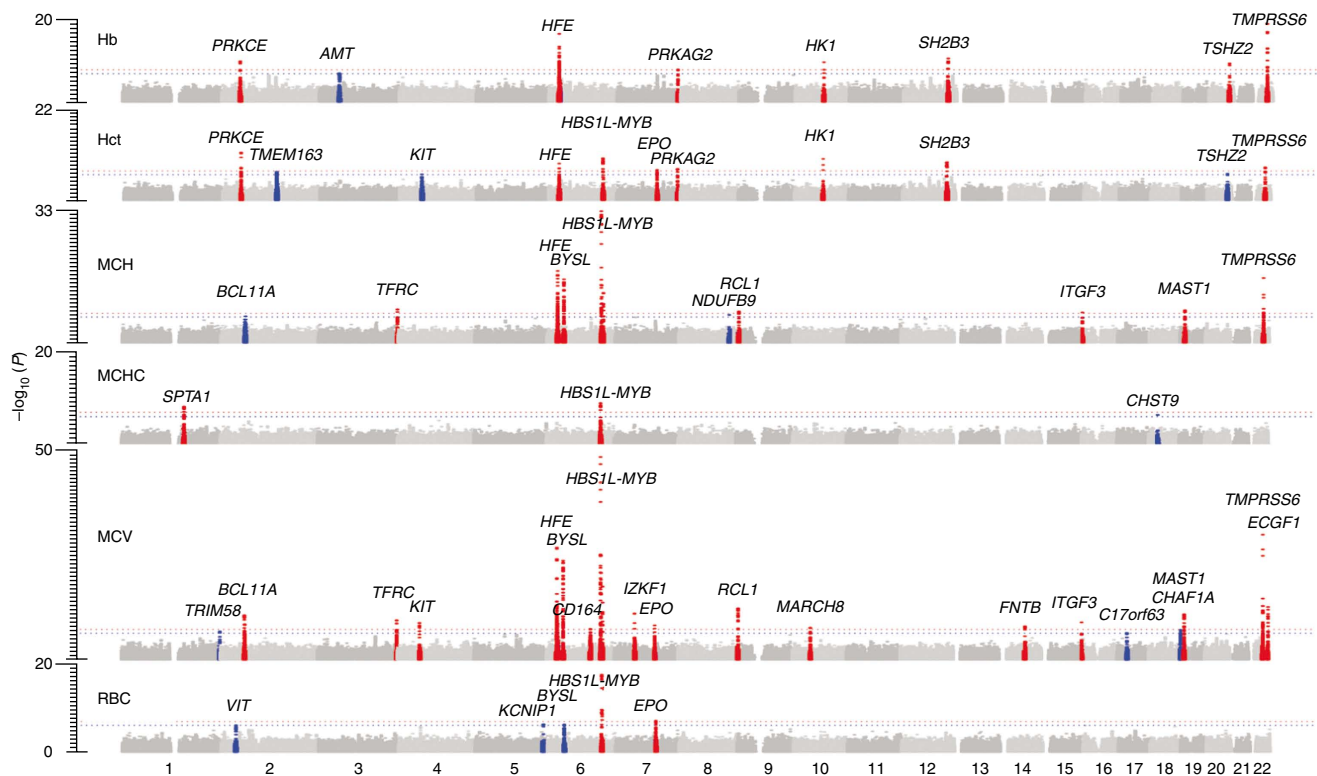
Sample sizes and summary statistics of covariates and erythrocyte traits measured in each cohort in CHARGE. Erythrocyte traits are abbreviated as follows: Hb, hemoglobin concentration; Hct, hematocrit; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; RBC, erythrocyte count. 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 3 s.d. of the population mean for each erythrocyte trait. Values for age and each erythrocyte trait are presented as mean (s.d.). NA, data not available.

## RESULTS

### Meta-analysis of GWAS for erythrocyte traits

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 an Italian cohort study, the Invecchiare in Chianti Study (InCHIANTI;  $N = 1,021$ ), in these analyses. Characteristics of the study participants, including age, sex and trait summaries, are presented in Table 1.

We studied six erythrocyte traits: hemoglobin concentration (Hb), 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 significance threshold of  $P < 5 \times 10^{-8}$ . The  $-\log_{10}(P)$  value genome-wide association plots for the meta-analysis of each of the six traits are shown in **Figure 1**, and corresponding quantile-quantile plots are shown in **Supplementary Figure 1a**. To assess population structure, we examined the per-cohort genomic control inflation factor,  $\lambda_{GC}$ , demonstrating that these values



**Figure 1** Overview of CHARGE meta-analysis results for six erythrocyte traits: hemoglobin concentration (Hb), hematocrit (Hct), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV) and erythrocyte count (RBC).  $-\log_{10}(P)$  value is plotted on the y axis against the 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 ( $P < 4 \times 10^{-7}$ ).

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

Locus no.	Chr	Number of SNPs per trait						In RefGene	Gene annotation	Closest RefGene
		Hb	Hct	MCH	MCHC	MCV	RBC			
1	1q23.1	0	0	0	8	0	0	<i>SPTA1</i>	<i>OR10Z1; SPTA1; OR10X1; OR6Y1</i>	
2	2p21	3	6	0	0	0	0	<i>PRKCE</i>	<i>PRKCE</i>	
3	2p16.1	0	0	0	0	27	0	<i>BCL11A</i>	<i>BCL11A</i>	
4	3q29	0	0	4	0	11	0	<i>TFRC</i>	<i>TFRC</i>	
5	4q12	0	0	0	0	8	0	None	None	<i>KIT</i>
6	6p22.2	49	3	72	0	133	0	<i>HFE; LRRC16; SCGN; SLC17A1; SLC17A3; SLC17A2; TRIM38; HIST1H4B; ZNF322A</i>	<i>HFE; SCGN; LRRC16; SLC17A1; SLC17A2; SLC17A3; SLC17A4; ZNF322A; ABT1; TRIM38</i>	
7	6p21.1	0	0	51	0	65	0	<i>PRICKLE4; FRS3; TFEB; MED20; USP49; CCND3; BYSL</i>	<i>PRICKLE4; FRS3; PGC; TFEB; MED20; USP49; CCND3; BYSL; TAF8; PGC; FRS3</i>	
8	6q21	0	0	0	0	5	0	None	<i>CD164</i>	
9	6q23.3	0	14	43	13	83	24	<i>HBS1L; MYB</i>	<i>ALDH8A1; HBS1L; MYB</i>	
10	6q24.1	0	0	9	0	13	0	None	None	<i>CITED2</i>
11	7p12.2	0	0	0	0	5	0	<i>IKZF1</i>	<i>IKZF1</i>	
12	7q22.1	0	2	0	0	2	2	<i>TFR2; ZAN</i>	<i>GNB2; PCOLCE; FBXO24; TFR2; ACTL6B; HRBL; MOSPD3; LRCH4; ZAN; EPO; POP7; PERQ1; EPHB4</i>	
13	7q36.1	1	3	0	0	0	0	<i>PRKAG2</i>	<i>PRKAG2</i>	
14	9p24.1	0	0	9	0	19	0	<i>RCL1</i>	<i>RCL1; AK3</i>	
15	10q11.21	0	0	0	0	4	0	<i>MARCH8</i>	<i>MARCH8; ALOX5</i>	
16	10q21.3	1	3	0	0	0	0	<i>HK1</i>	<i>HK1</i>	
17	12q24.12	10	9	0	0	0	0	<i>SH2B3; ATXN2; c12orf30; PTPN11</i>	<i>SH2B3; ATXN2; BRAP; ACAD10; ERP29; TMEM116; C12orf30; TRAFD1; C12orf30; RPL6; PTPN11</i>	
18	14q23.3	0	0	0	0	9	0	<i>MAX; FNTB</i>	<i>MAX; RAB15; FNTB</i>	
19	16p13.3	0	0	1	0	1	0	<i>ITFG3</i>	<i>LUC7L; PDIA2; AXIN1; ITFG3; RGS11; ARHGDIG</i>	
20	19p13.13	0	0	6	0	25	0	<i>MAN2B1; RTBDN; MAST1; DNASE2; GCDH; FARSa</i>	<i>MAN2B1; MORG1; ZNF490; TNPO2; DHPS; FBXW9; ZNF791; C19orf56; JUNB; HOOK2; PRDX2; RNASE-H2A; RTBDN; GADD45GIP1; KLF1; FARSa; RAD23A; CALR; MAST1; GCDH; DNASE2; DAND5; NFIX</i>	
21	20q13.2	3	0	0	0	0	0	None	None	<i>TSHZ2</i>
22	22q12.3	13	2	9	0	36	0	<i>C22orf33; TST; MPST; TMPRSS6</i>	<i>IL2RB; C22orf33; KCTD17; TST; TMPRSS6; MPST</i>	
23	22q13.33	0	0	0	0	12	0	<i>TMEM112B; NCAPH2; SCO2; ECGF1</i>	<i>TMEM112B; ADM2; MIOX; ECGF1; KLHDC7B; LOC440836; SAPS2; SCO2; NCAPH2; SBF1; MAP-K8IP2; MIOX; CHKB; CPT1B</i>	

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

were consistently below 1.08 (**Supplementary Table 2**). The genomic control inflation factor post meta-analysis, which was not corrected at the meta-analysis level, showed no systematic inflation (Hb  $\lambda_{GC} = 1.066$ ; 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 significance threshold. Altogether, there were 45 trait-locus combinations with at least one genome-wide 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.

### Replication studies

We pursued replication of the 45 SNPs identified from the initial meta-analysis in the CHARGE datasets, with a meta-analysis of association data in 9,456 independent European-ancestry individuals from five population-based cohorts of the HaemGen Consortium (**Supplementary Note**). A combined meta-analysis of the HaemGen and CHARGE data showed lower  $P$  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 Hb 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, and the distribution of replication  $P$  values is shown in **Supplementary Figure 1b**. 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 datasets.

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 Hb variation (7 SNPs); 1.16% of Hct variation (8 SNPs); 4.53% of MCH variation (9 SNPs); 0.63% of MCHC variation (2 SNPs); 5.98% of MCV variation (17 SNPs); and 0.85% of RBC variation (2 SNPs).

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

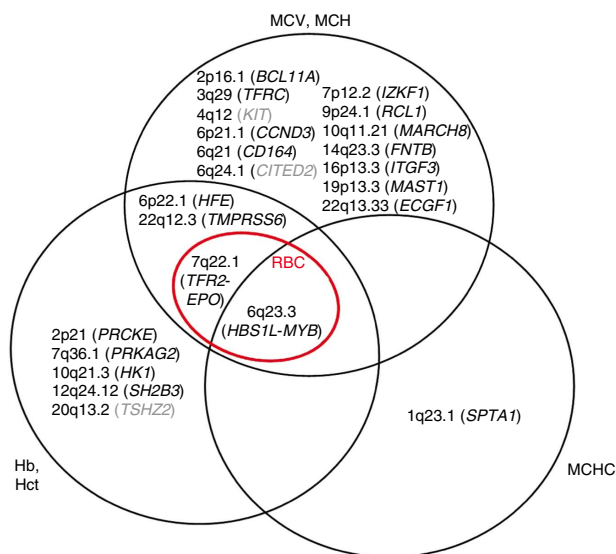
Locus no.	Trait	SNPID	Chr	PhysPos	min_all	MAF	Gene	CHARGE				HaemGen			CHARGE + HaemGen		
								% var	Beta	s.e.m.	<i>P</i>	Beta	s.e.m.	<i>P</i>	Beta	s.e.m.	<i>P</i>
2	Hb	rs10495928	2	46206670	G	0.341667	<i>PRKCE</i>	0.10%	0.011	0.011	$5.926 \times 10^{-10}$	0.0537	0.0152	<b>0.000414</b>	0.0631	0.0088	$7.052 \times 10^{-13}$
6	Hb	rs1800562	6	26201120	A	0.041667	<i>HFE</i>	0.17%	0.0224	0.0224	$3.592 \times 10^{-16}$	0.1167	0.0322	<b>0.000295</b>	0.1617	0.0182	$5.737 \times 10^{-19}$
13	Hb	rs10224002	7	151045974	G	0.258333	<i>PRKAG2</i>	0.14%	0.011	0.011	$4.605 \times 10^{-08}$	0.0964	0.0164	<b><math>4.19 \times 10^{-09}</math></b>	0.0714	0.0091	$3.025 \times 10^{-15}$
16	Hb	rs16926246	10	70763398	T	0.110169	<i>HK1</i>	0.21%	0.0192	0.0192	$1.036 \times 10^{-09}$	0.0863	0.0329	0.008721	0.1099	0.0164	$2.116 \times 10^{-11}$
17	Hb	rs11065987	12	111076069	A	0.333333	<i>TRAFD1</i>	0.16%	0.0107	0.0107	$1.345 \times 10^{-10}$	0.0427	0.0151	0.004591	0.0585	0.0086	$1.159 \times 10^{-11}$
21	Hb	rs6013509	20	50751758	A	0.183673	<i>TSHZ2</i>	0.15%	-0.0699	0.0116	$1.963 \times 10^{-09}$	-0.0479	0.0206	0.01999	-0.0646	0.01	$1.054 \times 10^{-10}$
22	Hb	rs855791	22	35792882	A	0.391667	<i>TMPPRSS6</i>	0.20%	-0.0962	0.011	$2.044 \times 10^{-18}$	-0.0845	0.0157	<b><math>7.97 \times 10^{-08}</math></b>	-0.0923	0.0089	$3.25 \times 10^{-25}$
2	Hct	rs10168349	2	46208555	C	0.341667	<i>PRKCE</i>	0.16%	0.201	0.0283	$1.176 \times 10^{-12}$	0.1524	0.0471	0.001212	0.1875	0.0238	$3.748 \times 10^{-15}$
6	Hct	rs1800562	6	26201120	A	0.041667	<i>HFE</i>	0.09%	0.3747	0.0608	$7.204 \times 10^{-10}$	0.1167	0.1004	0.245	0.3073	0.0513	$2.035 \times 10^{-09}$
9	Hct	rs9483788	6	135477194	C	0.177966	<i>HBS1L-MYB</i>	0.13%	0.2172	0.0328	$3.551 \times 10^{-11}$	0.2147	0.0527	<b><math>4.55 \times 10^{-05}</math></b>	0.2166	0.0274	$2.811 \times 10^{-15}$
12	Hct	rs7385804	7	100073906	C	0.377358	<i>TFR2</i>	0.17%	-0.1592	0.0286	$2.745 \times 10^{-08}$	-0.1269	0.0478	0.00799	-0.151	0.0242	$4.45 \times 10^{-10}$
13	Hct	rs10224002	7	151045974	G	0.258333	<i>PRKAG2</i>	0.22%	0.1691	0.0299	$1.492 \times 10^{-08}$	0.2727	0.0493	<b><math>3.08 \times 10^{-08}</math></b>	0.1963	0.0252	$6.045 \times 10^{-15}$
16	Hct	rs16926246	10	70763398	T	0.110169	<i>HK1</i>	0.13%	0.337	0.0513	$4.986 \times 10^{-11}$	0.3094	0.096	0.00127	0.3315	0.0445	$9.636 \times 10^{-14}$
17	Hct	rs11065987	12	110556807	G	0.341667	<i>SH2B3-ATXN2</i>	0.15%	-0.1809	0.0288	$3.343 \times 10^{-10}$	-0.1438	0.0465	0.001983	-0.171	0.0241	$1.363 \times 10^{-12}$
22	Hct	rs2413450	22	35800170	T	0.381818	<i>TMPPRSS6</i>	0.10%	-0.162	0.0279	$6.333 \times 10^{-09}$	-0.2082	0.0467	<b><math>8.13 \times 10^{-06}</math></b>	-0.1736	0.0236	$1.846 \times 10^{-13}$
4	MCH	rs11915082	3	197293536	A	0.425	<i>TFR2</i>	0.24%	0.0041	0.0007	$4.888 \times 10^{-09}$	0.0035	0.0009	<b><math>5.66 \times 10^{-05}</math></b>	0.0038	0.0005	$7.729 \times 10^{-13}$
6	MCH	rs1408272	6	25950930	G	0.033898	<i>SLC17A3</i>	0.41%	-0.0134	0.0015	$1.369 \times 10^{-18}$	-0.0183	0.0019	<b><math>2.37 \times 10^{-22}</math></b>	-0.0153	0.0012	$3.868 \times 10^{-39}$
7	MCH	rs9349205	6	42033137	A	0.196429	<i>CCND3/BYSL</i>	0.29%	-0.0066	0.0008	$1.785 \times 10^{-16}$	-0.0037	0.0009	<b><math>2.11 \times 10^{-05}</math></b>	-0.0053	0.0006	$8.198 \times 10^{-20}$
9	MCH	rs7776054	6	135460609	G	0.220339	<i>HBS1L/MYB</i>	1.02%	-0.0092	0.0008	$1.976 \times 10^{-33}$	-0.0107	0.0009	<b><math>3.08 \times 10^{-36}</math></b>	-0.0099	0.0006	$7.356 \times 10^{-69}$
10	MCH	rs628751	6	139880112	C	0.491667	<i>CITED2</i>	0.34%	-0.0049	0.0007	$3.84 \times 10^{-13}$	-0.0034	0.0008	<b><math>7.24 \times 10^{-06}</math></b>	-0.0043	0.0005	$1.262 \times 10^{-17}$
14	MCH	rs10758658	9	4846877	A	0.186441	<i>RCL1</i>	0.18%	-0.0048	0.0008	$1.634 \times 10^{-08}$	-0.0048	0.0009	<b><math>5.14 \times 10^{-07}</math></b>	-0.0048	0.0006	$2.166 \times 10^{-14}$
19	MCH	rs1122794	16	249156	A	0.224138	<i>ITFG3</i>	0.28%	0.0054	0.001	$2.992 \times 10^{-08}$	0.0036	0.0012	0.00299	0.0047	0.0007	$2.675 \times 10^{-10}$
20	MCH	rs11085824	19	12862547	G	0.366667	<i>GCDH</i>	0.20%	-0.0041	0.0007	$8.105 \times 10^{-09}$	-0.003	0.0009	<b>0.000532</b>	-0.0037	0.0005	$1.415 \times 10^{-11}$
22	MCH	rs2413450	22	35800170	T	0.381818	<i>TMPPRSS6</i>	0.41%	-0.006	0.0007	$8.818 \times 10^{-17}$	-0.0068	0.0008	<b><math>5.34 \times 10^{-18}</math></b>	-0.0064	0.0005	$8.77 \times 10^{-34}$
1	MCHCrs	rs857721	1	156879172	A	0.316667	<i>SPTA1</i>	0.33%	-0.0022	0.0004	$3.414 \times 10^{-09}$	-0.0013	0.0004	0.00266	-0.0018	0.0003	$1.033 \times 10^{-10}$
9	MCHCrs	rs9373124	6	135464902	C	0.220339	<i>HBS1L-MYB</i>	0.30%	-0.0023	0.0004	$6.486 \times 10^{-10}$	-0.0018	0.0004	<b><math>2.6 \times 10^{-05}</math></b>	-0.0021	0.0003	$7.003 \times 10^{-14}$
3	MCV	rs2540917	2	60462263	C	0.433333	<i>BCL11A</i>	0.24%	-0.0031	0.0005	$2.127 \times 10^{-11}$	-0.0022	0.0006	<b>0.000192</b>	-0.0028	0.0004	$1.125 \times 10^{-14}$
4	MCV	rs9859260	3	197284944	C	0.35	<i>TFR2</i>	0.23%	0.003	0.0005	$3.246 \times 10^{-10}$	0.003	0.0008	<b>0.000166</b>	0.003	0.0004	$8.499 \times 10^{-14}$
5	MCV	rs172629	4	55102519	G	0.116667	<i>KIT</i>	0.27%	-0.0043	0.0007	$1.36 \times 10^{-09}$	-0.0051	0.001	<b><math>2.22 \times 10^{-07}</math></b>	-0.0046	0.0006	$9.816 \times 10^{-16}$
6	MCV	rs1800562	6	26201120	A	0.041667	<i>HFE</i>	0.58%	0.0115	0.0011	$1.425 \times 10^{-27}$	0.0137	0.0015	<b><math>5.02 \times 10^{-20}</math></b>	0.0122	0.0009	$1.012 \times 10^{-46}$
7	MCV	rs9349205	6	42033137	A	0.196429	<i>CCND3-BYSL</i>	0.58%	-0.0055	0.0005	$1.756 \times 10^{-24}$	-0.0043	0.0008	<b><math>8.48 \times 10^{-08}</math></b>	-0.0051	0.0004	$1.121 \times 10^{-31}$
8	MCV	rs9374080	6	109723113	C	0.375	<i>CD164</i>	0.22%	-0.0026	0.0005	$3.695 \times 10^{-08}$	-0.0017	0.0006	0.003738	-0.0023	0.0004	$4.198 \times 10^{-10}$
9	MCV	rs4895441	6	135468266	G	0.225	<i>HBS1L-MYB</i>	1.12%	-0.008	0.0005	$1.004 \times 10^{-57}$	-0.0083	0.0008	<b><math>3.03 \times 10^{-27}</math></b>	-0.0081	0.0004	$7.241 \times 10^{-86}$
10	MCV	rs643381	6	139881116	A	0.489362	<i>CITED2</i>	0.50%	-0.0039	0.0005	$2.663 \times 10^{-18}$	-0.0037	0.0007	<b><math>1.63 \times 10^{-07}</math></b>	-0.0039	0.0004	$4.665 \times 10^{-25}$
11	MCV	rs12718597	7	50395922	A	0.275	<i>IKZF1</i>	0.26%	0.0032	0.0005	$8.138 \times 10^{-12}$	0.0018	0.0007	0.0145	0.0028	0.0004	$4.689 \times 10^{-13}$
12	MCV	rs7786877	7	100051951	G	0.208333	<i>TFR2</i>	0.13%	0.0032	0.0005	$5.452 \times 10^{-09}$	0.0024	0.0008	0.002081	0.003	0.0004	$2.543 \times 10^{-11}$
14	MCV	rs10758658	9	4846877	A	0.186441	<i>RCL1</i>	0.29%	-0.0041	0.0006	$4.354 \times 10^{-13}$	-0.0045	0.0008	<b><math>4.65 \times 10^{-08}</math></b>	-0.0043	0.0005	$3.184 \times 10^{-20}$
15	MCV	rs11239550	10	45344735	G	0.228814	<i>MARCH8</i>	0.15%	-0.0028	0.0005	$1.873 \times 10^{-08}$	-0.0022	0.0007	0.003496	-0.0026	0.0004	$1.346 \times 10^{-10}$
18	MCV	rs4466998	14	64545293	A	0.478723	<i>FNTB</i>	0.17%	0.0027	0.0005	$8.925 \times 10^{-09}$	0.0008	0.0006	0.2061	0.002	0.0004	$4.907 \times 10^{-08}$
19	MCV	rs7189020	16	244804	T	0.431034	<i>ITFG3</i>	0.19%	-0.0031	0.0005	$1.081 \times 10^{-09}$	-0.0024	0.0007	<b>0.000871</b>	-0.0029	0.0004	$1.819 \times 10^{-12}$
20	MCV	rs7255045	19	12793269	A	0.25	<i>RTBDN</i>	0.27%	-0.0037	0.0006	$1.233 \times 10^{-11}$	-0.0018	0.0008	0.0266	-0.0032	0.0004	$2.173 \times 10^{-12}$
22	MCV	rs2413450	22	35800170	T	0.381818	<i>TMPPRSS6</i>	0.65%	-0.0054	0.0005	$1.078 \times 10^{-30}$	-0.0046	0.0007	<b><math>7.62 \times 10^{-11}</math></b>	-0.0052	0.0004	$2.772 \times 10^{-41}$
23	MCV	rs131794	22	49318618	A	0.166667	<i>ECGF1</i>	0.22%	-0.0044	0.0006	$2.189 \times 10^{-13}$	-0.0029	0.0009	0.001795	-0.004	0.0005	$1.033 \times 10^{-15}$
9	RBC	rs9483788	6	135461324	G	0.225	<i>HBS1L-MYB</i>	0.65%	0.0141	0.0016	$3.115 \times 10^{-19}$	0.0155	0.0013	<b><math>2.19 \times 10^{-30}</math></b>	0.0141	0.001	$1.148 \times 10^{-47}$
12	RBC	rs2075671	7	100183042	A	0.225	<i>EPO</i>	0.20%	0.0086	0.0016	$3.058 \times 10^{-08}$	0.0047	0.0016	0.003383	0.0068	0.0011	$1.123 \times 10^{-09}$

Replication test results for the lead SNP per locus and per erythrocyte trait (45 SNPs). Results are organized by trait, then by locus, as indicated in **Table 1**. Results from a combined CHARGE and HaemGen Consortium meta-analysis are presented. SNPID, SNP Identifier; min\_all, minor allele; PhysPos, physical position. Minor allele frequency (MAF) is presented based on HapMap CEU. *P* values in bold font meet a Bonferroni-corrected significance threshold for replication of  $P < 0.0011$  (0.05/45). Units were: Hb, g/dl; Hct, %; MCH, pg; MCHC, g/dl; MCV, fl; RBC, 1 million cells/mm<sup>3</sup>.

### Annotation of associated loci

For 20 of the identified loci, top associated SNPs were identified within a window of  $\pm 60$  kb of a RefSeq gene (**Table 2**). For three loci, chromosomes 4q12, 6q24.1 and 20q13.2, no genes were identified within this window, with the nearest genes being 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*, *TMPPRSS6*, *SPTA1*, *HBS1L* and *MYB* (two closely adjacent genes), 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 nonhematologic diseases, and previously defined roles in hematologic and cardiovascular systems, are listed in **Supplementary Table 4**. We confirmed the association of the known mutations in the *HFE* gene resulting in the C282Y (rs1800562) and H63D (rs1799945)



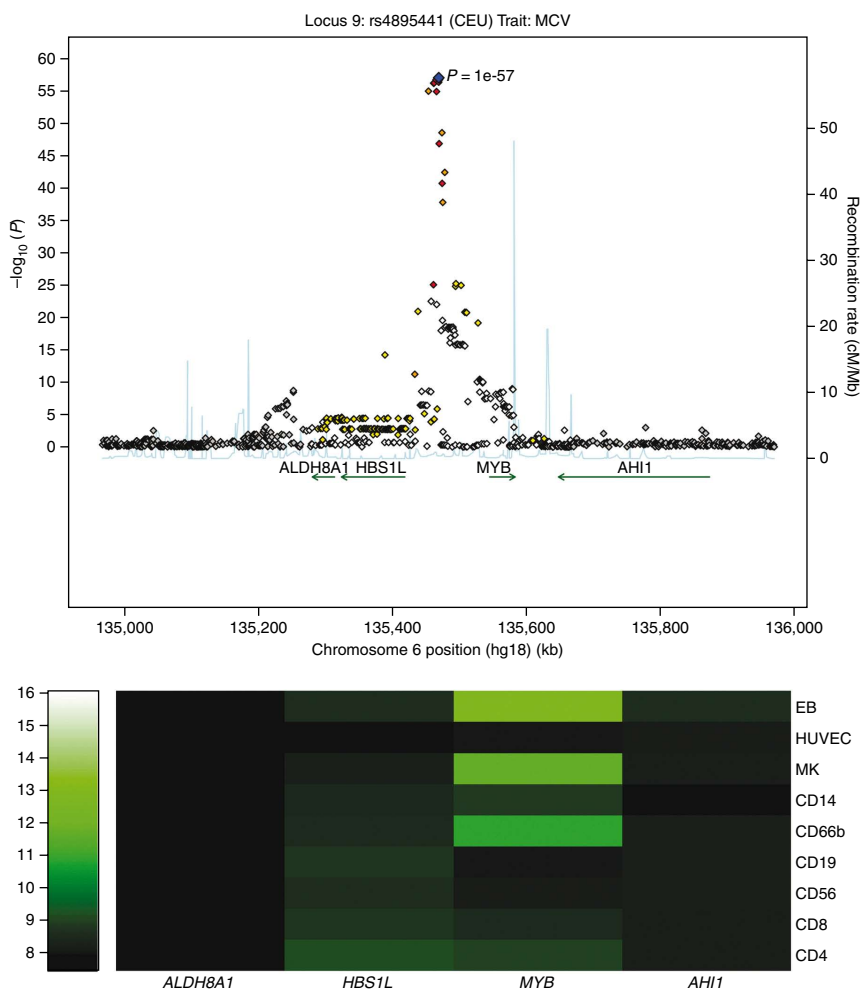
**Figure 2** Results of the CHARGE meta-analysis are organized into a Venn diagram, demonstrating the overlap of loci meeting a genome-wide significance threshold of  $P < 5 \times 10^{-8}$ .

substitutions, mutations that are already known to underlie hereditary hemochromatosis, with Hb, 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 erythroblasts (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*, *HBS1L*, *MYB*, *AHI1*) within the 1-Mb interval. A heat map showing gene expression levels for each of these four genes is shown in **Figure 3**, demonstrating approximately twofold expression of *MYB* in EBs as compared to other cell lines. Gene expression

**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  $\pm 500$  kb around the lead SNP identified in this locus, which is plotted in blue. SNPs identified by the CHARGE meta-analyses are color coded according to correlation with the lead SNP (red,  $r^2 \geq 0.8$ ; orange,  $0.5 \leq r^2 < 0.8$ ; yellow,  $0.2 \leq r^2 < 0.5$ ; white,  $r^2 < 0.2$ ; gray, no  $r^2$  value provided). The  $P$  value for the lead SNP in this region is provided. (b) A heat map of gene expression levels in nine blood and endothelial cell lines is shown, including all genes, as annotated by ENSEMBL 54, within the  $\pm 500$ -kb window of the locus. MK, megakaryocyte; EB, erythroblasts; HUVEC, human umbilical vein endothelial cells; CD14, monocytes; CD66b, granulocytes; CD19, B lymphocytes; CD56, NK cells; CD8, cytotoxic (Tc) T lymphocytes; CD4, T helper (Th) lymphocytes.



was detected for genes in multiple loci (**Supplementary Fig. 2**). Because 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 on chromosomes 3q29 (*TFRC*), 6p22.2 (*HIST1H4C*, which is near *HFE*), 6p21.1 (*CCND3*), 10q21.3 (*HK1*), 16p13.3 (*HBZ*, *HBA1*), and 22q12.3 (*TST*, *RAC2*). The genes most highly expressed in HUVECs were on chromosomes 6p22.2 (*HIST1H4C*), 7q22.1 (*SERPINE1*) and 22q13.3 (*MFNG*).

### Association with blood pressure

The results of association testing for the 45 lead SNPs within 23 loci, as identified from the CHARGE meta-analysis, are summarized in **Supplementary Table 5a**. We identified associations at a Bonferroni-corrected significance threshold of  $P < 0.00135$  ( $0.05/37$ , as 37 of the 45 SNPs are unique) on chromosomes 12q24.1 (*SH2B3*) and 7q36.1 (*PRKAG2*). The previously reported association of the chromosome 12q24.1 locus *SH2B3* 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 blood pressure and HTN analysis, signals at this locus spanned 700 kb, from rs3184504 to rs11066188, and association signals in Hb and Hct spanned 987 kb, from rs3184504 in *SH2B3* to rs11066301 in *PTPN11*, and contained multiple genes<sup>15</sup>. Inspection of RNA expression data (**Supplementary Fig. 2**) showed that among

genes in this region, *SH2B3* and *ATXN2* are highly expressed in erythroblasts 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**); results of the evaluation of SNPs associated with blood pressure and hypertension are presented in **Supplementary Table 5b**.

## DISCUSSION

In this meta-analysis of genome-wide association data from 24,167 individuals of European ancestry from six cohort studies in the CHARGE Consortium, we identified 23 loci associated with at least one of the six erythrocyte traits Hb, Hct, MCH, MCHC, MCV and RBC. We sought evidence for replication in an additional 9,456 individuals of European ancestry from cohorts of the HaemGen Consortium. In the joint meta-analysis, combining CHARGE and HaemGen datasets, all 23 loci had  $P$  values  $< 5 \times 10^{-8}$ , meeting our genome-wide significance threshold and suggesting that they merit further study. Among the 23 loci, six were previously known quantitative trait loci (QTLs), and 17 are newly associated loci, some of which include candidate genes whose products are 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, although none of these met our genome-wide significance threshold.

The six erythrocyte traits examined here include some that are highly correlated, and as expected, we observed a high degree of concordance in the results across the six traits. Notably, MCHC, a ratio of Hb 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<sup>16,17</sup>. Reviewing the overall association results, we observed there were generally three patterns of significant associations among the six traits (**Fig. 2**). Results were generally similar for (i) Hb and Hct, which are mainly quantitative measures of hemoglobin in the blood; (ii) MCH and MCV, representing erythrocyte size and quantity of hemoglobin per erythrocyte; and (iii) MCHC, the ratio of Hb to Hct, which appears somewhat distinct from the other traits. Across the six traits studied, the strongest signal was found in the *HBS1L-MYB* locus on chromosome 6q23, which was observed for five of the six individual traits (Hct, MCH, MCHC, MCV and RBC) at the genome-wide significance level. This locus also provided a modest but nonsignificant result for the sixth trait, Hb (rs4895441,  $P = 4.8 \times 10^{-4}$ ). In addition, the Hb/Hct and MCH/MCV patterns overlapped for associations to 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 Hb/Hct and MCH/MCV patterns, with associations observed in the 7q22.1 (*TFR2-EPO*) and 6q23 (*HBS1L-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.

We identified genome-wide significant association of SNPs within the *HFE* gene with Hb, Hct, MCH and MCV. The mutation in *HFE*

encoding the C282Y substitution is the principal cause of hereditary hemochromatosis, a common autosomal recessive iron overload disease in individuals of northern European descent<sup>18</sup>. This mutation was associated with increased MCV and Hb concentrations in a study of individuals drawn from a screening study for hemochromatosis and iron overload<sup>7</sup>, and this variant was the lead association result for both Hb and Hct in our study. Individuals heterozygous for either allele do not manifest clinical iron overload but may have an increased iron uptake and resistance to anemia, and the C282Y-encoding mutation may increase the risk of coronary heart disease by increasing iron stores and lipid oxidation<sup>19,20</sup>. 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 through its actions in both downregulating the absorption of iron in the intestine and inhibiting the release of iron from macrophages<sup>21–24</sup>. SNPs within *TMPRSS6* were associated with Hb, Hct, MCH and MCV. *TMPRSS6* was identified by linkage and association studies in five families and two sporadically affected individuals with iron-refractory iron-deficiency anemia, a rare mendelian disease<sup>25</sup>. *TMPRSS6* encodes a type II transmembrane serine protease produced by the liver that regulates the expression of hepcidin<sup>25</sup>. The transferrin receptor (encoded by *TFR1*, also called *TFR1*) and transferrin receptor 2 (*TFR2*) are highly homologous type II transmembrane proteins in the transferrin protein family. SNPs within *TFR1* were associated with MCH and MCV, and SNPs within *TFR2* were associated with Hct and MCV. Reduced *TFR1* expression is associated with anemia<sup>26</sup>. Existing evidence indicates that *TFR2* is also a modulator of hepcidin expression and that mutations in *TFR2* cause hemochromatosis type 3 (ref. 27).

Two loci, chromosome 2p16.1 (*BCL11A*) associated with MCV, and chromosome 6q23.3 (*HBS1L-MYB*) associated with all traits except for Hb, are related to variation in fetal hemoglobin and hemoglobin- $\beta$  levels<sup>28,29</sup>. *BCL11A* is an oncogene related to B-cell malignancies<sup>30</sup> and regulates fetal hemoglobin expression<sup>31</sup>. *BCL11A* is expressed in erythroid precursors, and we observed *BCL11A* expression in EBs (**Supplementary Fig. 2**), making it a biologically plausible candidate gene for erythrocyte trait variation<sup>32</sup>. A healthy-population study showed that polymorphisms in *HBS1L* and *MYB* influences erythrocyte, platelet and monocyte counts<sup>10</sup>. Although the role of *HBS1L* is unknown, *MYB* has been associated with proliferation, survival and differentiation of hematopoietic progenitor cells<sup>33,34</sup>. *MYB* is also associated with eosinophil counts in blood and atopic asthma<sup>35</sup>. There are strong associations between SNPs within this locus and multiple erythrocyte traits (lead SNP rs4895441, Hb  $P = 4.8 \times 10^{-4}$ ; 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 earlier linkage analyses of erythrocyte traits have identified this chromosomal region<sup>6,11,13</sup>. SNPs within *SH2B3* are associated with Hct and Hb. Notably, the SNPs within this gene are associated with blood pressure, myocardial infarction, type 1 diabetes and celiac disease<sup>15,35–39</sup>. *SH2B3* is expressed in hematopoietic precursor cells and increases hematopoietic progenitors of erythroid, megakaryocytic and myeloid lineages<sup>35,40</sup>. *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 effects 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*<sup>-/-</sup> model<sup>41</sup>, and *PTPN11* mutations cause Noonan's and LEOPARD syndromes and juvenile myelomonocytic leukemia<sup>42–44</sup>. Lastly, we identified associations for

Hct, MCV and RBC near the *EPO* gene. Erythropoietin, the *EPO* gene product, 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 complications<sup>45</sup> but not with erythrocyte traits.

SNPs within the *SPTA1* gene were associated with MCHC. *SPTA1* encodes erythroid spectrin, a protein in the erythrocyte membrane that 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<sup>16,17</sup>.

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 of our rationale for studying the relationship of the SNPs identified through the study of erythrocyte traits to blood pressure and hypertension. Earlier studies have shown that Hb and Hct levels are associated with increased risks for hypertension and a variety of other vascular diseases and for mortality<sup>1–3,46–49</sup>. From a rheologic perspective, blood viscosity depends largely on Hb or Hct levels and is a determinant of blood pressure<sup>50–53</sup>. There is an inverse relationship between viscosity and vascular blood flow<sup>54</sup>, and elevated Hct thereby hampers organ perfusion. Given these findings, we are intrigued by the overlap between the association results for Hb and Hct from our study and the recently reported associations for blood pressure and hypertension<sup>15,36</sup>. We observed overlap of associations in the chromosome 12q24.12 region across a 98-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*; this gene does not have specifically known hematologic or vascular roles, but mutations in the gene cause cardiomyopathy and disorders of the cardiac conduction system<sup>55,56</sup>. Neither causality nor independence of these associations is necessarily supported by these findings. However, the associations suggest that common genetic bases may underlie some of the correlation seen between erythrocyte, blood pressure and hypertension traits. Further confirmation in large independent cohorts may provide stronger evidence for the strength and consistency of the associations with hypertension.

One limitation of the current analyses is that our study cohorts were limited to individuals of European ancestry. Several mutations affecting spectrin and globin 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 ancestry<sup>57</sup>. 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. In the replication analysis, for those loci that did not meet a conservative replication test, power may have been limited, and many of these associations 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 regard to erythrocyte production, homeostasis and function. Specific differences in association patterns between the individual traits examined may highlight different pathways, but further studies are needed to characterize these.

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 blood pressure and hypertension. Further studies are needed to examine these associations in ethnically diverse populations and to characterize the functional impact of variants at the implicated candidate genes.

## METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturegenetics/>.

*Note: Supplementary information is available on the Nature Genetics website.*

## ACKNOWLEDGMENTS

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

## AUTHOR CONTRIBUTIONS

*AGES*: Study design and phenotype collection, T.B.H., V.G., L.J.L., G.E.; data analysis, A.V.S., M.A.N., T.A.; manuscript preparation, M.A.N., N.S.; manuscript revisions, T.B.H., V.G., L.J.L., A.V.S., A.B.S., D.G.H., M.A.N. *ARIC*: Study design and phenotype collection, E.B., J.C., S.F., A.C.; genotyping, E.B., A.C.; data analysis, S.K.G., A.K., A.C., A.G., G.B.E., A.J.; manuscript preparation, S.K.G., S.F.; manuscript revisions, S.K.G., A.C., E.B., J.C., A.K., S.F. *CHS*: Study design and phenotype collection, B.M.P., J.I.R., M.C., N.A.Z., N.L.G.; genotyping, J.I.R.; data analysis, N.L.G., T.L.; manuscript preparation, N.A.Z.; manuscript revisions, B.M.P., J.C.M.B., K.R., J.I.R., M.C., N.A.Z., N.L.G., T.L. *FHS*: Study design and phenotype collection, M.-H.C., Q.Y., C.S.F., D.L., L.A.C., C.J.O., J.-P.L.; data analysis, M.-H.C., Q.Y., J.-P.L., G.Z.; manuscript preparation, M.-H.C., C.J.O., J.-P.L.; manuscript revisions, M.-H.C., Q.Y., C.S.F., D.L., L.A.C., C.J.O., J.-P.L. *Rotterdam*: Study design and phenotype collection, F.J.A.v.R., A.H., A.G.U., B.A.O., C.M.v.D., J.C.M.W.; genotyping, A.G.U.; data analysis, F.J.A.v.R., J.F.F., A.D., G.C.V.; manuscript preparation, F.J.A.v.R., J.F.F., C.M.v.D.; manuscript revisions, F.J.A.v.R., J.F.F., A.D., G.C.V., A.H., A.G.U., B.A.O., C.M.v.D., J.C.M.W. *InCHIANTI*: Study design and phenotype collection, L.F., J.M.G., S.B.; data analysis, M.A.N., T.T.; manuscript preparation, M.A.N.; manuscript revisions, L.F., J.M.G., S.B., K.V.P., A.B.S., D.G.H., M.A.N., T.T. *HaemGen*: A.R., J.-J.Z., J.M.v.G. *Twins UK*: Study design and phenotype collection, T.D.S., S.-L.T., P.D.; data analysis and manuscript preparation and revisions, N.S. *UKBS1*: Study design and phenotype collection, W.H.O., J.G.S.; data analysis, N.S.; manuscript preparation and revisions, N.S., A.R. *KORA*: Study design and phenotype collection, C.M., C.G.; data analysis, C.G., B.K.; manuscript preparation and revisions, C.G. *SHIP*: Study design and phenotype collection, M.N., A.G.; data analysis, A.T.; gene expression, A.R., J.-J.Z., J.M.v.G.

## COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturegenetics/>.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Havlik, R.J. *et al.* Evidence for additional blood pressure correlates in adults 20–56 years old. *Circulation* **61**, 710–715 (1980).
- Tell, G.S. *et al.* Correlates of blood pressure in community-dwelling older adults. The Cardiovascular Health Study. Cardiovascular Health Study (CHS) Collaborative Research Group. *Hypertension* **23**, 59–67 (1994).
- Zakai, N.A. *et al.* A prospective study of anemia status, hemoglobin concentration, and mortality in an elderly cohort: the Cardiovascular Health Study. *Arch. Intern. Med.* **165**, 2214–2220 (2005).
- Whitfield, J.B. & Martin, N.G. Genetic and environmental influences on the size and number of cells in the blood. *Genet. Epidemiol.* **2**, 133–144 (1985).
- Evans, D.M., Frazer, I.H. & Martin, N.G. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res.* **2**, 250–257 (1999).
- Lin, J.P. *et al.* Evidence for linkage of red blood cell size and count: genome-wide scans in the Framingham Heart Study. *Am. J. Hematol.* **82**, 605–610 (2007).
- McLaren, C.E. *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.* **82**, 898–905 (2007).
- Zeng, S.M., Yankowitz, J., Widness, J.A. & Strauss, R.G. Etiology of differences in hematocrit between males and females: sequence-based polymorphisms in erythropoietin and its receptor. *J. Genet. Specif. Med.* **4**, 35–40 (2001).

9. Zeng, S.M., Yankowitz, J., Widness, J.A. & Strauss, R.G. Sequence-based polymorphisms in members of the apoptosis Bcl-2 gene family and their association with hematocrit level. *J. Genet. Specif. Med.* **6**, 36–42 (2003).
10. Menzel, S. *et al.* The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. *Blood* **110**, 3624–3626 (2007).
11. Iliadou, A. *et al.* Genome-wide scans of red cell indices suggest linkage on chromosome 6q23. *J. Med. Genet.* **44**, 24–30 (2007).
12. Yang, Q., Kathiresan, S., Lin, J.P., Tofler, G.H. & O'Donnell, C.J. Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. *BMC Med. Genet.* **8** (Suppl. 1), S12 (2007).
13. Lin, J.P., O'Donnell, C.J., Levy, D. & Cupples, L.A. Evidence for a gene influencing haematocrit on chromosome 6q23–24: genome-wide scan in the Framingham Heart Study. *J. Med. Genet.* **42**, 75–79 (2005).
14. Psaty, B.M. *et al.* 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. Cardiovasc. Genet.* **2**, 73–80 (2009).
15. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**, 677–687 (2009).
16. Agre, P., Casella, J.F., Zinkham, W.H., McMillan, C. & Bennett, V. Partial deficiency of erythrocyte spectrin in hereditary spherocytosis. *Nature* **314**, 380–383 (1985).
17. Marchesi, S.L. *et al.* Mutant forms of spectrin alpha-subunits in hereditary elliptocytosis. *J. Clin. Invest.* **80**, 191–198 (1987).
18. Feder, J.N. *et al.* A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat. Genet.* **13**, 399–408 (1996).
19. Roest, M. *et al.* Heterozygosity for a hereditary hemochromatosis gene is associated with cardiovascular death in women. *Circulation* **100**, 1268–1273 (1999).
20. Tuomainen, T.P. *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* **100**, 1274–1279 (1999).
21. Vecchi, C. *et al.* ER stress controls iron metabolism through induction of hepcidin. *Science* **325**, 877–880 (2009).
22. Swinkels, D.W. & Wetzels, J.F. Hepcidin: a new tool in the management of anaemia in patients with chronic kidney disease? *Nephrol. Dial. Transplant.* **23**, 2450–2453 (2008).
23. Roy, C.N. *et al.* Hepcidin antimicrobial peptide transgenic mice exhibit features of the anemia of inflammation. *Blood* **109**, 4038–4044 (2007).
24. Feder, J.N. *et al.* The hemochromatosis gene product complexes with the transferrin receptor and lowers its affinity for ligand binding. *Proc. Natl. Acad. Sci. USA* **95**, 1472–1477 (1998).
25. Finberg, K.E. *et al.* Mutations in Tmprss6 cause iron-refractory iron deficiency anemia (IRIDA). *Nat. Genet.* **40**, 569–571 (2008).
26. Kerenyi, M.A. *et al.* Stat5 regulates cellular iron uptake of erythroid cells via IRP-2 and TfR-1. *Blood* **112**, 3878–3888 (2008).
27. Roetto, A. *et al.* Hemochromatosis due to mutations in transferrin receptor 2. *Blood Cells Mol. Dis.* **29**, 465–470 (2002).
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. USA* **105**, 1620–1625 (2008).
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.* **39**, 1197–1199 (2007).
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* **106**, 3183–3190 (2005).
31. Sankaran, V.G. *et al.* Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322**, 1839–1842 (2008).
32. Higgs, D.R. & Wood, W.G. Genetic complexity in sickle cell disease. *Proc. Natl. Acad. Sci. USA* **105**, 11595–11596 (2008).
33. Sandberg, M.L. *et al.* c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Dev. Cell* **8**, 153–166 (2005).
34. Emambokus, N. *et al.* Progression through key stages of haemopoiesis is dependent on distinct threshold levels of c-Myb. *EMBO J.* **22**, 4478–4488 (2003).
35. Gudbjartsson, D.F. *et al.* Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.* **41**, 342–347 (2009).
36. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).
37. Todd, J.A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* **39**, 857–864 (2007).
38. Hunt, K.A. *et al.* Newly identified genetic risk variants for celiac disease related to the immune response. *Nat. Genet.* **40**, 395–402 (2008).
39. Smyth, D.J. *et al.* Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N. Engl. J. Med.* **359**, 2767–2777 (2008).
40. Velazquez, L. *et al.* Cytokine signaling and hematopoietic homeostasis are disrupted in Lnk-deficient mice. *J. Exp. Med.* **195**, 1599–1611 (2002).
41. Saxton, T.M. *et al.* Abnormal mesoderm patterning in mouse embryos mutant for the SH2 tyrosine phosphatase Shp-2. *EMBO J.* **16**, 2352–2364 (1997).
42. Jongmans, M. *et al.* Genotypic and phenotypic characterization of Noonan syndrome: new data and review of the literature. *Am. J. Med. Genet. A* **134A**, 165–170 (2005).
43. Kalidas, K. *et al.* Genetic heterogeneity in LEOPARD syndrome: two families with no mutations in PTPN11. *J. Hum. Genet.* **50**, 21–25 (2005).
44. Tartaglia, M. *et al.* Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat. Genet.* **34**, 148–150 (2003).
45. Tong, Z. *et al.* Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications. *Proc. Natl. Acad. Sci. USA* **105**, 6998–7003 (2008).
46. Wannamethee, G. & Shaper, A.G. Haematocrit: relationships with blood lipids, blood pressure and other cardiovascular risk factors. *Thromb. Haemost.* **72**, 58–64 (1994).
47. Gagnon, D.R., Zhang, T.J., Brand, F.N. & Kannel, W.B. Hematocrit and the risk of cardiovascular disease—the Framingham study: a 34-year follow-up. *Am. Heart J.* **127**, 674–682 (1994).
48. Elwood, P.C., Waters, W.E., Benjamin, I.T. & Sweetnam, P.M. Mortality and anaemia in women. *Lancet* **1**, 891–894 (1974).
49. Sarnak, M.J. *et al.* Anemia as a risk factor for cardiovascular disease in The Atherosclerosis Risk in Communities (ARIC) study. *J. Am. Coll. Cardiol.* **40**, 27–33 (2002).
50. de Simone, G. *et al.* Association of blood pressure with blood viscosity in american indians: the Strong Heart Study. *Hypertension* **45**, 625–630 (2005).
51. Letcher, R.L., Chien, S., Pickering, T.G. & Laragh, J.H. Elevated blood viscosity in patients with borderline essential hypertension. *Hypertension* **5**, 757–762 (1983).
52. Sharp, D.S. *et al.* Mean red cell volume as a correlate of blood pressure. *Circulation* **93**, 1677–1684 (1996).
53. Wells, R.E. Jr. & Merrill, E.W. Influence of flow properties of blood upon viscosity-hematocrit relationships. *J. Clin. Invest.* **41**, 1591–1598 (1962).
54. Ott, E.O., Lechner, H. & Aranibar, A. High blood viscosity syndrome in cerebral infarction. *Stroke* **5**, 330–333 (1974).
55. Gollob, M.H. *et al.* Identification of a gene responsible for familial Wolff-Parkinson-White syndrome. *N. Engl. J. Med.* **344**, 1823–1831 (2001).
56. Gollob, M.H. *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* **104**, 3030–3033 (2001).
57. Beutler, E. & Waalen, E. The definition of anemia: what is the lower limit of normal of the blood hemoglobin concentration? *Blood* **107**, 1747–1750 (2006).

<sup>1</sup>National Human Genome Research Institute, Division of Intramural Research, Bethesda, Maryland, USA. <sup>2</sup>National Heart, Lung, and Blood Institute (NHLBI), Division of Intramural Research, Bethesda, Maryland, USA. <sup>3</sup>McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, USA. <sup>4</sup>Departments of Medicine and Pathology, University of Vermont, Burlington, Vermont, USA. <sup>5</sup>Department of Epidemiology, Erasmus Medical Center (MC), Rotterdam, The Netherlands. <sup>6</sup>The Netherlands Genomics Initiative-sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA), Leiden, The Netherlands. <sup>7</sup>Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK. <sup>8</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. <sup>9</sup>Icelandic Heart Association, Kopavogur, Iceland. <sup>10</sup>Laboratory of Neurogenetics, National Institute of Aging (NIA), National Institutes of Health (NIH), Bethesda, Maryland, USA. <sup>11</sup>NHLBI's Framingham Heart Study, Framingham, Massachusetts, USA. <sup>12</sup>Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA. <sup>13</sup>Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA. <sup>14</sup>Department of Medicine, University of Washington, Seattle, Washington, USA. <sup>15</sup>Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, USA. <sup>16</sup>Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. <sup>17</sup>University of Iceland, Reykjavik, Iceland. <sup>18</sup>Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. <sup>19</sup>Longitudinal Studies Section—Clinical Research Branch, NIA, NIH, Baltimore, Maryland, USA. <sup>20</sup>MedStar Research Institute, Hyattsville, Maryland, USA. <sup>21</sup>Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>22</sup>Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-Universität Greifswald, Greifswald, Germany. <sup>23</sup>Laboratory for Epidemiology, Demography, and Biometry, NIA, NIH, Bethesda, Maryland, USA. <sup>24</sup>Cardiology Center, Geneva University Hospital, Geneva, Switzerland. <sup>25</sup>Department of Biostatistics, University of Washington, Seattle, Washington, USA. <sup>26</sup>Department of Haematology, University of Cambridge and National Health Service Blood and Transplant, Long Road, Cambridge, UK. <sup>27</sup>Human Genetics Center, University of Texas Health Science Center, Houston, Texas, USA. <sup>28</sup>Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. <sup>29</sup>Intramural Research Program, Laboratory of Neurogenetics, NIA, NIH, Bethesda, Maryland, USA. <sup>30</sup>Office of Biostatistical Research, Division of Cardiovascular Sciences, NHLBI, NIH, Bethesda, Maryland, USA. <sup>31</sup>Geriatric Unit ASF, Firenze, Italy. <sup>32</sup>Department of Clinical Chemistry, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>33</sup>Leiden University Medical Center, Leiden, The Netherlands. <sup>34</sup>Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands. <sup>35</sup>Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands. <sup>36</sup>Department of Experimental Immunohematology, Sanquin Research, Amsterdam, The Netherlands. <sup>37</sup>Department of Haematology, King's College London, London, UK. <sup>38</sup>Institut für Klinische Chemie und Laboratoriumsmedizin, Ernst-Moritz-Arndt Universität Greifswald, Greifswald, Germany. <sup>39</sup>Department of Epidemiology, University of Washington, Seattle, Washington, USA. <sup>40</sup>Department of Health Services, University of Washington, Seattle, Washington, USA. <sup>41</sup>Group Health Research Institute, Group Health, Seattle, Washington, USA. <sup>42</sup>Institut für Immunologie und Transfusionsmedizin, Ernst-Moritz-Arndt-Universität Greifswald, Greifswald, Germany. <sup>43</sup>Cardiology Division, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. <sup>44</sup>Department of Pediatrics, Johns Hopkins School of Medicine, Baltimore, Maryland, USA. <sup>45</sup>These authors contributed equally to this work. Correspondence should be addressed to S.K.G. (ganeshs@mail.nih.gov) or C.J.O. (odonnellc@mail.nih.gov).

## ONLINE METHODS

**CHARGE Consortium organization and study samples.** We performed a cross-sectional analysis of genotype and phenotypic data on erythrocyte traits in the CHARGE Consortium<sup>14</sup>, 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 and CHS, Illumina 550K; InCHIANTI, RS). Genotypes were then imputed to a set of approximately 2.5 million HapMap SNPs using Phase II CEU HapMap individuals for reference using either MACH (ARIC, AGES, FHS, InCHIANTI, RS) or BAMBAM<sup>58</sup> (CHS) software.

**Erythrocyte phenotypes.** Erythrocyte parameters studied were (i) hemoglobin concentration (Hb), the concentration of hemoglobin within whole blood; (ii) hematocrit (Hct), the percentage of whole blood comprising cellular erythrocyte elements; (iii) red blood cell count (RBC), the number of red blood cells per volume of blood; (iv) mean corpuscular volume (MCV), the average erythrocyte volume; (v) mean corpuscular hemoglobin (MCH), the average quantity of hemoglobin per erythrocyte; and (vi) mean corpuscular hemoglobin concentration (MCHC), the ratio of Hb to Hct. The definition and units for 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 Hb and Hct were not transformed before analyses. In order to focus on determinants of variation of these traits in the general population rather than on specific hematologic diseases, which are overrepresented at the tails of the distribution for each of the traits, we restricted analysis to those individuals within 3 s.d. of the sample mean within each cohort. For each SNP meeting quality control 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 multicenter cohorts. In FHS, linear mixed effects models were used to account for relatedness, and these models included adjustment for principal components, computed using Eigenstrat 2.0 (ref. 59). The genome-wide level of significance threshold was set at  $P < 5 \times 10^{-8}$ .

The result for each cohort, including regression coefficients, standard errors, sample size, imputation quality, minor allele designation and minor allele frequency, were forwarded to a central repository. Before meta-analysis, we performed genomic control on each cohort-specific distribution of the association test statistics for each trait<sup>60</sup>. We also filtered out SNPs with allelic frequency <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. After meta-analysis, the genomic control inflation factor,  $\lambda_{GC}$ , 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 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 lead 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,  $N = 1,510$ ; Kooperative Gesundheitsforschung in der Region Augsburg (KORA)-F3 500 K study population,  $N = 1,643$ ; and KORA-F4 1000 K study population,  $N = 1,814$ ). Further information on these cohorts is provided in the **Supplementary Note**. Trait definitions were identical to those for 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 the impact of increasing sample size 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 EBs and megakaryocytes (MKs), CD14<sup>+</sup> monocytes, CD56<sup>+</sup> natural killer (NK) cells, CD4<sup>+</sup> T helper (Th) lymphocytes, CD8<sup>+</sup> cytotoxic T (Tc) lymphocytes and CD66b<sup>+</sup> granulocytes) using an established catalog of gene expression in these lineages<sup>61</sup>; we also examined gene expression in cultured HUVECs<sup>62</sup>. For each of the 23 identified loci, we examined the transcript levels in the nine cell types of all genes within a  $\pm 500$ -kb window centered on the lead association SNP.

Briefly, EB and MK samples were obtained by culturing CD34<sup>+</sup> hematopoietic stem cells purified from cord blood cultured, respectively, for 10 d with erythropoietin, interleukin-3 and stem cell factor (EPO-IL3-SCF) and for 7 d with thrombopoietin (TPO) and interleukin-1 $\beta$  (TPO-IL1 $\beta$ )<sup>61</sup>. Cultured cells were sorted by fluorescence-activated cell sorting using a monoclonal antibody either against CD235a (glycophorin A) or against CD41 ( $\alpha_{IIb}$  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 bead arrays; the detailed methods for isolation of other hematologic cell types, RNA extraction and microarray analysis are described elsewhere<sup>61</sup>. Cultured HUVECs and expression profiles were ascertained as has been described elsewhere<sup>62</sup> and using the same microarray platform<sup>61</sup>.

Raw data was transformed with the variance-stabilizing transform (which optimally stabilizes the expression variance across the intensity range for Illumina bead arrays) using the R/Bioconductor package "lumi"<sup>63</sup> and then subjected to 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 <0.01 were removed (implemented in the R package "outliers"). This was not done for the MK, EB or HUVEC cell types due to their smaller sample sizes (4, 4 and 3, respectively). Probe mappings were obtained from reannotation efforts available online.

**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, OMIM<sup>64</sup> and the Genetic Association Database<sup>65</sup> (**Supplementary Table 4**).

**Blood pressure and hypertension analyses.** Following the observation that epidemiologic studies have identified a strong yet unexplained link between Hb 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<sup>1,2,15</sup>. The CHARGE cohorts that contributed samples to the blood pressure 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 trait<sup>15</sup>. We tested the association of the lead SNP per locus for each trait (45 SNPs) for association with SBP, DBP and hypertension<sup>15</sup>, 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 hypertension<sup>15</sup>.

**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/research/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

version 1.0.15/16 <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/>.

58. Servin, B. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.* **3**, e114 (2007).
59. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
60. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
61. Watkins, N.A. *et al.* A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood* **113**, e1–e9 (2009).
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.* **27**, 1023–1029 (2007).
63. Du, P., Kibbe, W.A. & Lin, S.M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
64. Hamosh, A (ed.). Online Mendelian Inheritance in Man (McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Maryland, USA) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Maryland, USA)). <<http://www.ncbi.nlm.nih.gov/omim/>>.
65. Becker, K.G., Barnes, K.C., Bright, T.J. & Wang, S.A. The genetic association database. *Nat. Genet.* **36**, 431–432 (2004).