Meta-analysis of rare and common exome chip variants identifies S1PR4 and other loci influencing blood cell traits

Nathan Pankratz, Ursula M Schick, Yi Zhou, Wei Zhou, Tarunveer Singh Ahluwalia, Maria Laura Allende, Paul L Auer, Jette Bork-Jensen, Jennifer A Brody, Ming-Huei Chen, Vinna Clavo, John D Eicher, Niels Grarup, Elliott J Hagedorn, Bella Hu, Kristina Hunker, Andrew D Johnson, Maarten Leusink, Yingchang Lu, Leo-Pekka Lyytikäinen, Ani Manichaikul, Riccardo E Marioni, Mike A Nalls, Raha Pazoki, Albert Vernon Smith, Frank A van Rooij, Min-Lee Yang, Xiaoling Zhang, Yan Zhang, Folkert W Asselbergs, Eric Boerwinkle, Ingrid Borecki, Erwin P Bottinger, Mary Cushman, Richard L Proia, Leonard I Zon, Christopher J O’Donnell, Alex P Reiner & Santhi K Ganesh for the CHARGE Consortium Hematology Working Group

Hematologic measures such as hematocrit and white blood cell (WBC) count are heritable and clinically relevant. We analyzed erythrocyte and WBC phenotypes in 52,531 individuals (37,775 of European ancestry, 11,589 African Americans, and 3,167 Hispanic Americans) from 16 population-based cohorts with Illumina HumanExome BeadChip genotypes. We then performed replication analyses of new discoveries in 18,018 European-American women and 5,261 Han Chinese. We identified and replicated four new erythrocyte trait–locus associations (CEP89, SHROOM3, FADS2, and APOE) and six new WBC loci for neutrophil count (S1PR4), monocyte count (BTBD8, NLRP12, and IL17RA), eosinophil count (IRF1), and total WBC count (MYB). The association of a rare missense variant in S1PR4 supports the role of sphingosine-1-phosphate signaling in leukocyte trafficking and circulating neutrophil counts. Loss-of-function experiments for S1PR4 in mouse and s1pr4 in zebrafish demonstrated phenotypes consistent with the association observed in humans and altered kinetics of neutrophil recruitment and resolution in response to tissue injury.

Erythrocyte and leukocyte blood counts are heritable traits (estimated heritability of 0.40–0.90 (refs. 1–3) and 0.14–0.40 (ref. 4), respectively) that reflect core physiological functions of oxygen-carrying capacity and antimicrobial activity. Peripheral blood cell counts are commonly measured in the clinical setting to diagnose and monitor therapy for many acute and chronic conditions, such as infection or anemia. Abnormalities in these clinical measures often reflect primary hematologic disease, blood loss, or inflammation. Between-individual differences in erythrocyte traits, total WBC counts, and neutrophil counts have been associated with risk of cardiovascular diseases and all-cause mortality.

Previous genome-wide association studies (GWAS) have defined over 100 loci influencing erythrocyte traits and leukocyte counts. However, few studies have systematically evaluated the contribution of coding variation, particularly variants at low frequency in the general population. Recently completed exome sequencing in diverse populations has led to international collaboration and creation of a genome-wide catalog of low-frequency coding variants. We undertook a large-scale study of erythrocyte and leukocyte traits in up to 52,531 individuals of European, African, and Hispanic ancestry to evaluate the impact of both low-frequency and common variants assayed by the Illumina HumanExome BeadChip, also referred to as the exome chip.

RESULTS

Study samples

In the discovery stage, we analyzed erythrocyte traits (hemoglobin concentration (Hb), hematocrit (Hct), mean corpuscular hemoglobin...
(MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), red blood cell distribution width (RDW), and red blood cell (RBC) count and leukocyte traits (total WBC count and absolute neutrophil, lymphocyte, monocyte, eosinophil, and basophil counts) (Supplementary Table 1) in 52,531 individuals, including 37,775 with European ancestry, 11,589 African Americans, and 3,167 white Hispanics sampled from 16 population-based cohorts. The sample sizes for each trait analyzed in each cohort are provided in Supplementary Table 2. The traits followed expected distributions, and the characteristics of the study participants, including age, sex, and trait summaries, are presented in Supplementary Table 3. Descriptions of each cohort are provided in the Supplementary Note.

Single-variant and gene-based meta-analyses

In single-variant analyses, we did not observe marked inflation of the meta-analysis P values (Supplementary Table 4). A total of 104 unique locus–trait associations exceeded the Bonferroni-corrected significance threshold (P < 4 × 10−7; Supplementary Table 5). The loci included 49 independent loci associated with erythrocyte traits and 22 independent loci associated with leukocyte traits (Supplementary Table 6). Many of these were SNPs well established to have associations with hematologic traits (Supplementary Note), thus confirming the validity of the exome chip. New findings reaching study-wide significance (P < 4 × 10−7; n = 9 loci for erythrocytes and n = 10 loci for leukocytes) are listed in Table 1 and were carried forward to replication in an independent sample. Of these findings, four new trait–locus associations for erythrocyte traits (SHROOM3, CEP89, and APOE) were study-wide significant, P < 0.003; FADS2 was only nominally significant, P = 0.02) and six new trait–locus associations for WBC traits (BTRD8, MYB–HBS1L1, SIPR4, and IL17RA) were study-wide significant, P < 0.003; IRF1 and NLRP12 were only nominally significant, P < 0.05) were replicated in an independent set of European-American samples from the Women’s Health Initiative (WHI; Supplementary Table 7).

New replicated associations with erythrocyte traits

All four new replicated erythrocyte associations are common SNPs present on the exome chip because of previous associations with non-hematologic phenotypes as listed in the National Human Genome Research Institute (NHGRI) GWAS catalog. Two common intronic SNPs previously associated with renal function (SHROOM3/r s13146355 and CEP89/rs4805834) were associated with both Hb and Hct. The SHROOM3/rs13146355 [A] minor allele (previously associated with both lower estimated glomerular filtration rate (eGFR) and higher serum magnesium levels (18) was associated with significantly higher Hb and Hct and with nominally higher RBC count in our discovery and replication cohorts. The CEP89/rs4805834[T] minor allele was associated with lower Hb and Hct and with higher eGFR (15). The observed directions of effect on Hb and Hct for both CEP89/rs4805834 and SHROOM3/rs13146355 are opposite to those expected on the basis of the known relationship between lower eGFR and anemia. Conditional analyses performed in a subset of our cohorts demonstrated that the effect of either CEP89/rs4805834 or SHROOM3/rs13146355 on Hb and Hct was independent of eGFR (see the Supplementary Note for more detail).

An intronic SNP of the fatty acid desaturase gene FADS2 (rs1535) previously associated with transferrin levels (20) and polyunsaturated fatty acid (PUFA) levels (21) was associated with RBC count. Finally, we identified an association between increased RDW and the minor allele of rs7412, encoding the canonical APOE ε2 variant, which is known to be associated with cholesterol (22,23) and inversely associated with dementia (24,25). Additional adjustment for LDL cholesterol, HDL cholesterol, and triglyceride levels did not attenuate the APOE ε2–RDW association in the Atherosclerosis Risk in Communities (ARIC) study. A tag for the APOE ε4 allele was present on the exome chip but had no association with either LDL cholesterol or RDW independent of APOE ε2 (Supplementary Note).

In gene-based tests, several loci were significantly associated with erythrocyte traits in the discovery sample (Table 2 and Supplementary Table 8). Bold font indicates a significant association in the discovery set (P < 5 × 10−7), a significant association in the replication set (P < 0.003), or a nominal association in the replication set (P < 0.05). Chr., chromosome; EA, European ancestry; AA, African American; HA, Hispanic American.
**Table 2. Top results for gene-based tests in the discovery and replication samples**

| Trait | Gene | Replication P | WHI EA | WHI HA | EA + AA + HA | EA AA | HA | EA + AA + HA | EA AA | HA | WHI EA | Hb, Hct, MCH, MCHC, MCV |
|-------|------|---------------|--------|--------|--------------|-------|----|--------------|-------|----|--------|-----------------------|
| Red blood cell traits | | | | | | | | | | | | | |
| Trait | Gene | Replication P | WHI EA | WHI HA | EA + AA + HA | EA AA | HA | EA + AA + HA | EA AA | HA | WHI EA | Hb, Hct, MCH, MCHC, MCV |
| | | | | | | | | | | | | | |

**Table 8.** The EPO gene-based association was driven by a single low-frequency missense variant (p.Asp70Asn/rs62483572), confirming the recent association of this variant with lower Hb. Similarly, a new variant (p.Asp534Asn/rs144091859) drove the FAM234A (better known as ITFG3) association. The HFE and G6PD associations were driven by population-specific common variants identified in previous GWAS that were included in the gene-based test because they are common in one population but absent in another and therefore averaged out to below the minor allele frequency (MAF) < 0.05 threshold for inclusion in the trans-ancestry analysis. Significant associations for ANK1, NLRC3, and HBS1L were supported by multiple rare variants (Supplementary Table 9a and Supplementary Note).

**New replicated leukocyte associations**

We discovered and replicated six new WBC trait–locus associations. In the single-variant analysis, we identified a single missense variant in the type 4 sphingosine-1-phosphate receptor (S1PR4), p.Arg365Leu/rs34436714, that was associated with lower total WBC counts (P = 3.4 × 10⁻⁷) and lower neutrophil counts (P = 3.4 × 10⁻⁷) (Supplementary Fig. 1). The association was consistent across cohorts (Fig. 1) and validated in both replication samples (WHI European-American women, P = 0.001; Peking University–University of Michigan Study of Atherosclerosis (PUUMA) Han Chinese, P = 0.003; Pmeta (discovery + replication) = 5 × 10⁻¹²) (Supplementary Table 7). The variant is rare (MAFmeta = 0.006) and not in linkage disequilibrium (LD) with variants in the region (Supplementary Fig. 1). In both the discovery and replication analyses, p.Arg365Leu was the only variant contributing to the significant gene-based association. Neutrophil counts were approximately 10% lower in carriers of the minor allele for p.Arg365Leu (Fig. 2). The p.Arg365Leu amino acid substitution is located in the intracellular cytoplasmic tail of S1PR4, is at a conserved site (GERP score of 3.94), and is predicted to be ‘possibly damaging’ by PolyPhen-2 (ref. 28).

Two missense variants were associated with lower monocyte count: a low-frequency p.Val601Leu variant in BTBD8 (rs34856868; European MAF = 0.03) and a common p.Gly39Val variant in NLRP12 (rs34436714; European MAF = 0.217). Three common intergenic variants included on the exome chip as GW AS index SNPs originally associated with non-leukocyte phenotypes were newly associated with WBC traits in our analysis. The common intergenic regulatory variant of HBS1L–MYB (previously associated with erythrocyte and platelet traits) was associated with total WBC count. Common non-coding SNPs in the regions of IL33 and IRF1 previously associated with asthma and other allergic and autoimmune disorders were associated with eosinophil count.

Gene-based analyses identified an association between low-frequency variation in the IL17RA locus and monocyte count (P = 6.4 × 10⁻⁷). We confirmed the recently reported multivariate association between CXCR2 and lower neutrophil count; six of the nine rare CXCR2 missense variants in our analysis had P < 0.05, with the strongest associations from p.Arg153His (rs55799208; P = 2.4 × 10⁻⁵) and p.Arg248Gln (rs61773609; P = 6.1 × 10⁻⁵). Several additional single-variant and gene-based associations with WBC traits were observed in the African-American or Hispanic-American discovery sample but not in the larger European ancestry discovery sample. Three of these associations were driven by low-frequency (MAF = 0.01–0.05) variants in African Americans (IQCJ) and SEC24D or Hispanic Americans (SEC14L1) (Table 1 and Supplementary Tables 8 and 9b). Further assessment in independent African-American or Hispanic-American samples will be needed to validate these ancestry-specific associations.
Characterization of variants in previously known GWAS loci

To evaluate whether variants identified in our analysis overlapped previously known GWAS results or whether we had identified independent associations, we conducted conditional analyses in the ARIC study, adjusting for previously known variants associated with erythrocyte and leucocyte traits in several regions overlapping the findings in this study (Supplementary Table 10). Specifically, we interrogated any variant that was rare (MAF <5%) and met study-wide significance (Supplementary Note).

We identified a new association between a low-frequency variant in ANKI (p.Ala1462Val/rs34664882; European MAF = 0.029, African-American MAF = 0.015, and Hispanic-American MAF = 0.013) and MCHC that was independent of the original GWAS result (rs4737009; 1000 Genomes Project CEU (European) MAF = 0.27 and ARIC European-American MAF = 0.24). We also identified several low-frequency missense variants in the HBA1–HBA2 region on chromosome 16. The most prominent was an African-American-specific variant in TTFG3 (p.Asp534Asn) that was associated with several erythrocyte parameters (MCH, MCHC, MCV, and RBC count) and is independent of the common GWAS association (Supplementary Note). Significant associations with the same traits were also seen for rare variants in MRPL28, NARFL, RGS11, TMEM8A, and TPSD1 (Supplementary Note).

Expression quantitative trait locus analysis

We used expression quantitative trait locus (eQTL) analysis54 to determine whether newly identified noncoding variants were associated with the expression of nearby genes across a range of tissue types (Supplementary Table 11). The most notable eQTL findings were in the FADS2 locus, which was associated with RBC count in our discovery analysis and reached a nominal significance level in the replication analysis (P = 0.02). In this region, FADS1, FADS2, and FADS3 all showed evidence of strong cis-eQTL association with the index SNP (rs1535) in multiple tissues, including for FADS1 (minimum P = 8.0 × 10^{-31} in CD19+ B cells) and FADS2 (minimum P = 3.0 × 10^{-57} in blood lymphocytes). The S1PR4 variant encoding p.Arg365Leu did not demonstrate an association with the expression levels of S1PR4 or any nearby transcript (Supplementary Table 12).

Among the new and independently replicated loci, rs4895441 in the HBS1L–MYB locus showed the expected eQTL association with HBS1L expression in multiple tissues (minimum P = 3.1 × 10^{-34} in aortic endothelial cells). In the SHROOM3 locus, rs131463 exhibited a weak eQTL association (P = 7.3 × 10^{-9}) with SHROOM3 transcript expression in subcutaneous adipose tissue. In the CEP89 locus, rs4805834 was associated with expression of SLC7A9 in multiple tissues (P = 1.9 × 10^{-24} in whole blood). The IRF1 SNP, rs12521868, was associated with expression of IRF1 in multiple tissues (P = 1.4 × 10^{-125} in whole blood).

Pleiotropy in the associated loci

In addition to pleiotropy of our new findings with known associations for kidney function (CEP89 and SHROOM3) and with dementia and dyslipidemia (APOE), we also identified variants with pleiotropy across multiple blood cell lineages, most notably for the HBS1L–MYB and SH2B3 loci, as well as other subthreshold associations (Supplementary Table 13 and Supplementary Note).

Confirmation of S1PR4 as a causal gene in model systems

The primary hypothesis of our exome chip study was that focused evaluation of coding variation would identify new genetic associations of rare variants with hematologic traits and that these variants would be more likely to be functionally relevant owing to the selection of variants for the exome chip. Our study yielded many new associations, in part because of coverage of noncoding variation included on the exome chip as follow-up for previous GWAS, as was the case for the four new RBC-associated loci we report here. The association of a rare missense variant in S1PR4 with total WBC and neutrophil counts was consistent with our a priori hypothesis, and we therefore undertook further follow-up studies of this gene’s functional impact on neutrophil traits in model systems.

Figure 2 Distributions of neutrophil counts for carriers and non-carriers of the S1PR4 variant encoding p.Arg365Leu in the ARIC study.
Using previously generated S1pr4-null mice,35 we evaluated peripheral circulating blood neutrophil and monocyte counts, bone marrow neutrophil counts, and spleen neutrophil counts in S1pr4−/− mice and S1pr4+/+ littermates. We analyzed 12 mice in each genotype group (total n = 24 mice), with equal numbers of males and females in each group, and found that the mean percentage of total cells analyzed by FACS that were double positive for Gr-1 and CD11b, marking neutrophils, was 31% lower in S1pr4−/− mice than in wild-type mice. We repeated the experiment in an additional 24 mice, again with 12 mice in each genotype group and equal numbers of males and females in each group, and saw a similar decrease. Across the 48 mice, both the percentage of WBCs that were neutrophils (28.0% decrease; P = 0.11) and the absolute neutrophil count (54.3% decrease; P = 0.03) were lower in S1pr4−/− mice than in wild-type mice (Fig. 3, Supplementary Fig. 2, and Supplementary Table 14). To evaluate the effects on circulating monocyte counts, FACS analysis was conducted in the same samples and the mean percentage of total leukocytes analyzed by FACS that were Gr-1+CD11b+ was equivalent in the two mouse groups (6.36% in S1pr4+/+ mice and 6.20% in S1pr4−/− mice; P = 0.80; Supplementary Figs. 3 and 4). Because abnormalities in leukocyte bone marrow egress have been described in the setting of S1pr1 deficiency,36 we evaluated the proportion of neutrophils in the bone marrow and spleen, to evaluate whether cells might be abnormally retained in these tissues, and the expression of specific adhesion molecules involved in leukocyte trafficking. No significant differences in neutrophil proportion or absolute counts were observed in the bone marrow or spleen for mice from the two genotype groups (Supplementary Figs. 3–6 and Supplementary Table 14). CD49d and CXCR4 were not differentially expressed on bone marrow neutrophils (P > 0.05), whereas the expression of CD62L, or l-selectin, measured on circulating neutrophils, was reduced by approximately twofold in S1pr4−/− mice (P = 0.003) across both groups of mice studied. Because lower l-selectin levels may reflect shedding upon activation and cell extravasation into tissues, we evaluated tissue neutrophil numbers in the liver and lung of S1pr4−/− and wild-type mice. Neutrophil numbers were lower in both tissues in S1pr4−/− mice, with a 29.2% reduction in liver (P = 0.12) and a 40.2% reduction in lung (P = 0.02) (Supplementary Fig. 7).

To further assess the impact of disrupted s1pr4 expression in vivo, we conducted parallel experiments in zebrafish, in which gene expression may be readily manipulated using morpholino oligonucleotide (MO) antisense technology to specifically knock down the expression of target genes.37 In the comparison of embryos injected with ATG-MOs designed against two independent sequences (Supplementary Note) in the single exon of s1pr4 (n = 14 and 19) to non-specific MO (n = 22), we confirmed a 36.6% and 34.3% decrease in neutrophil count in the two batches of whole embryos at 2 days post-fertilization (d.p.f.) (P = 3.8 × 10−6 and P = 4.4 × 10−7, respectively) (Fig. 4 and Supplementary Table 15).

Finally, to assess neutrophil behavior in response to injury, a cutaneous wound was made on the ventral side of the tail fin of...
the embryos at 2 d.p.f. after treatment with s1pr4 MO 1 in comparison to uninjected controls, and the numbers of neutrophils around the wound area at intervals up to 5 h after injury were counted to quantify neutrophil recruitment and resolution in response to the injury. The overall number of neutrophils recruited to the wound was higher and recruitment took place faster in embryos treated with s1pr4 MO; however, after initial recruitment, a trend toward a higher reverse migration rate and fewer cells retained at the site of injury was seen in the s1pr4 morphants (Supplementary Table 16), suggesting that the time course of neutrophil response to injury and resolution of inflammation may be altered in the setting of decreased s1pr4 expression (Fig. 5).

DISCUSSION
Using a custom genotyping array with focused coverage of missense and loss-of-function variants in exonic regions, we conducted an analysis of erythrocyte and leukocyte traits in as many as 52,531 individuals of European, African and Hispanic ancestry. We identified and replicated nine new genetic loci associated with total WBC count and erythrocyte and WBC phenotypes. These findings confirmed a role for this gene in two model organisms. Our findings highlight the importance of genes involved in erythrocyte membrane composition and leukocyte trafficking in the regulation of peripheral erythrocyte and WBC phenotypes.

The rare missense variant in S1PR4 (p.Arg365Leu/rs3746072; MAFmeta = 0.006) was robustly associated with total WBC count and neutrophil count. S1PR4 belongs to a family of G-protein-coupled protein receptors for spingosine-1-phosphate (S1P), a lysophospholipid that functions as an extracellular signaling molecule with diverse biological functions, including leukocyte trafficking. Another S1P receptor subtype, S1PR1, has an important role in regulating immune cell function and lymphocyte trafficking by regulating egress of lymphocytes from the bone marrow and lymphoid tissues; however, much less is known about the function of S1PR4. The S1PR4 receptor is expressed on hematopoietic and lymphoid cells and has been implicated in terminal megakaryocyte differentiation to platelets and in the regulation of dendritic cell function and T helper type 1 (Th1) cell13 and plasmacytoid dendritic cell differentiation. S1PR4 is highly expressed in neutrophils and lymphocytes. In the setting of combined Sgp1l (SIP lyase) and S1pr4 deletion in mice, neutrophilia and inflammation are decreased in comparison to SIP lyase deficiency alone. This suggests that S1pr4 may mediate the higher neutrophil count that accompanies highly elevated SIP levels in mice with SIP lyase deficiency.

Here we confirm in two in vivo vertebrate model systems (mouse and zebrafish) that loss of S1pr4 function leads to lower basal numbers (and proportion) of circulating neutrophils, consistent with the association observed in human carriers of the p.Arg365Leu variant.

The mild reduction in neutrophil count suggests a hypothesis of abnormal neutrophil trafficking, rather than a critical role for S1pr4 in neutrophil development. Bone marrow egress of leukocytes is known to be impaired in the setting of S1pr1 deficiency in mice. We therefore examined the expression of previously defined key adhesion molecules for leukocyte migration in response to S1P signaling in the bone marrow neutrophils of S1pr4-null mice, including CD49d, which is abnormally expressed in S1pr1- and SIP lyase-deficient states46, and CXCR4, which interacts with the cytokine peptide SDF-1 required for cellular bone marrow egress46.47. We did not find any alterations in the levels of these adhesion molecules and did not observe accumulation of neutrophils in the bone marrow or spleen, corroborating a lack of effect for S1pr4 on egress of neutrophils. Because neutrophil recruitment to injured or infected tissue is a key process, we evaluated the neutrophil cell surface expression of CD62L, which mediates interactions between neutrophils and endothelium and is required for leukocyte trafficking across the endothelial border. Neutrophil expression of CD62L was reduced by approximately twofold in S1pr4-null mice. Because CD62L is shed from the surface of neutrophils upon activation and mediates leukocyte extravasation into tissues, we examined whether tissue neutrophil counts were elevated in S1pr4-null mice, thereby accounting for lower circulating neutrophil counts. However, tissue neutrophil numbers were not increased in S1pr4-null mice in comparison to wild-type mice; rather, they were similarly decreased as in the blood.

In zebrafish s1pr4 morphants, neutrophil accumulation and resolution at the site of a cutaneous wound occurred earlier than in controls, suggesting impaired cellular inflammation in response to tissue injury in the absence of s1pr4. Further experiments to delineate neutrophil–endothelial cell interactions will be needed to further delineate the precise mechanisms by which S1PR4 influences circulating neutrophil counts. Together, our observations support the role of S1PR4 in the regulation of neutrophil counts and potentially clinically relevant impairment in response to injury or infection.
Blood monocyte counts are altered in the setting of chronic inflammatory disease and various infections, both viral and non-viral. NLRP12 attenuates inflammation by suppressing nuclear factor (NF)-κB signaling in activated monocytes\textsuperscript{49}. Loss-of-function mutations in NLRP12 have been identified in families with hereditary periodic fever syndromes\textsuperscript{50}. The missense variant reported here, p.Gly39Val, was not reported in these families and is not present in the ClinVar database\textsuperscript{51}. In humans and mice, NLRP12 is highly expressed in bone marrow and macrophages from Nlrp12-deficient mice exhibit decreased chemotaxis in response to chemokines in vitro, together suggesting that NLRP12 is important for leukocyte cell trafficking\textsuperscript{52}. IL17RA encodes a proinflammatory cytokine with a role in hematopoietic cell maturation, and vascular IL-17RA supports monocyte adherence\textsuperscript{53,54}. Mutations in IL17RA are associated with familial candidiasis\textsuperscript{55}.

Eosinophil counts are altered in parasitic infection and in allergic and autoimmune diseases such as asthma and inflammatory bowel disease. The IL33 variant rs1342326 has previously been associated with asthma\textsuperscript{29}, and IL-33 activates eosinophils\textsuperscript{56}. The IRF1 variant rs12521868, which we show to be associated with IRF1 expression, has previously been associated with Crohn’s disease\textsuperscript{57}. IRF1 is also near IL5, a known regulator of eosinophil production previously associated with eosinophil count\textsuperscript{58}.

We observed new erythrocyte trait associations for common non-coding SNPs representing two genomic regions previously associated with kidney function\textsuperscript{19,59}, near SHROOM3 and near CEP89. For both loci, the allele associated with lower Hb and Hct was associated with higher eGFR\textsuperscript{17}, suggesting that these erythrocyte trait associations are not mediated through an effect on renal dysfunction and related decreases in erythropoietin production. The SHROOM3 locus has additionally been associated with serum magnesium levels\textsuperscript{18}. The effects of these two loci on erythrocyte, renal, and electrolyte traits may instead be mediated by cytoskeleton-dependent solute or ion channels shared by kidney epithelium and erythrocyte membranes, as has been demonstrated for other loci such as PIEZO1, another GWAS-identified locus for erythrocyte traits. PIEZO1 encodes an RBC membrane mechanosensitive cation channel that appears to require actin cytoskeleton reorganization for activation\textsuperscript{60,61} and senses mechanical forces associated with fluid flow and/or circumferential stretch in epithelial cells at the basolateral side of renal proximal convoluted tubules\textsuperscript{62–64}. Dominant missense mutations in PIEZO1 have been reported in patients with hereditary xerocytosis\textsuperscript{65}, a congenital hemolytic anemia characterized by dehydrated, shrunken erythrocytes and the presence of stomatocytes due to increased potassium permeability. By analogy, SHROOM3 encodes an actin-binding protein involved in epithelial shape regulation, modulating ion channel activity through myosin II–dependent cytoskeletal reorganization in the kidney\textsuperscript{66}. Its role in erythrocyte function has yet to be tested experimentally. CEP89 is a ubiquitously expressed and highly conserved gene for which biological function is not well known. CEP89 is flanked by SLC7A9, which encodes a kidney solute transporter. Mutations in SLC7A9 result in congenital cystinuria\textsuperscript{67}. Our eQTL analyses showed a significant association of the index SNP associated with Hb and Hct with SLC7A9 transcript levels in multiple tissue types and most strongly in whole blood, supporting a possible hematologic function for SLC7A9.

FADS1 and FADS2 encode the two rate-limiting desaturases in the conversion of dietary essential medium-chain PUFA\textsubscript{s} (for example, α-linoleic acid (ALA)) into long-chain PUFA\textsubscript{s} (arachidonic acid, eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA)). The rs1535[G] minor allele of the FADS2 intronic variant is associated with higher levels of ALA and lower levels of EPA and DHA, as well as cholesterol levels\textsuperscript{68}. This suggests less efficient conversion due to decreased FADS activity\textsuperscript{21}. Here we report that the same FADS2 allele is associated with higher RBC count, Hb, and Hct. rs1535 is in strong LD with other common SNPs in the FADS1–FADS2 region on chromosome 11q12.2, including several eQTL SNPs for FADS1. Our eQTL analysis of this region showed strong associations of these SNPs with FADS1, FADS2, and FADS3 expression levels. Long-chain PUFA\textsubscript{s} are incorporated into erythrocyte membrane glycerolipids, affecting erythrocyte membrane fluidity, permeability, and sensitivity to oxidative damage and subsequent hemolysis\textsuperscript{69}. Nonetheless, the association of rs1535 with higher RBC count suggests additional mechanisms. In this regard, rs1535 is also located ~100 kb from FHIT, which encodes the heavy subunit of ferritin, the major intracellular iron storage protein that is expressed in both mature erythrocytes and early erythroid precursors.

Pleiotropy (more than one trait associated with the same locus) was observed for erythrocyte associations at the CEP89, FADS1, and HFE loci, and we extended the association of the well-characterized common HBSL1–MYB regulatory variant, previously associated with erythrocyte and platelet traits, to WBC count. MYB encodes c-Myb, a transcription factor and proto-oncoprotein expressed in immature hematopoietic cells and leukemic cells that has an essential role in the regulation of normal hematopoiesis and leukemogenesis\textsuperscript{70}. In addition, we confirmed the previously reported association of the chromosome 12q24 SH2B3 region with erythrocyte and WBC traits; this locus has been associated with multiple cardiovascular and inflammatory traits and diseases\textsuperscript{6,71–73}.

Our results add to recent observations that rare coding variants contribute to phenotypic differences in complex blood cell traits among community-dwelling individuals unscreened for hematologic disorders. Experimental testing of loss of S1pr4 function in vivo, performed to follow up a rare missense variant association in S1PR4 in our study, identified new biological effects on neutrophil count and function. Common variants originally associated with a single blood cell trait through GWAS, such as variants in the SH2B3 region, have subsequently been associated with traits related to all three blood cell lineages\textsuperscript{10,74}, as well as non-hematologic traits\textsuperscript{73,75,76}, and these pleiotropic effects will be useful to discern patterns suggesting specific biological hypotheses for further testing of mechanisms.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** Summary data for all analyses have been deposited in the database of Genotypes and Phenotypes (dbGaP) under CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) Consortium Summary Results from Genomic Studies. The dbGaP study accession is phs000930.

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank the staff and participants of all studies for their important contributions. A complete list of acknowledgments for each study is available in the Supplementary Note.

This work was supported by the following grants and contracts: US National Institutes of Health contracts (N01AG12100, HHSN26820110005C,
43. Schulze, T. et al. Sphingosine-1-phosphate receptor 4 (S1P) deficiency profoundly affects dendritic cell function and T<sub>17</sub>-cell differentiation in a murine model. *FASEB J.* **25**, 4024–4036 (2011).
44. Dillmann, C., Mora, J., Olesch, C., Brüne, B. & Weigert, A. S1PR4 is required for plasmacytid dendritic cell differentiation. *Biol. Chem.* **396**, 775–785 (2015).
45. Oliveira, A. et al. Sphingosine kinase 1 and sphingosine-1-phosphate receptor 2 are vital to recovery from anaphylactic shock in mice. *J. Clin. Invest.* **120**, 1429–1440 (2010).
46. Eash, K.J., Greenbaum, A.M., Gokalp, P.K. & Link, D.C. CXCR2 and CXCR4 antagonistically regulate neutrophil trafficking from murine bone marrow. *J. Clin. Invest.* **120**, 2423–2431 (2010).
47. Beck, T.C., Gomes, A.C., Cyster, J.G. & Pereira, J.P. CXCR4 and a cell-extrinsic mechanism control immature B lymphocyte egress from bone marrow. *J. Exp. Med.* **211**, 257–268 (2012).
48. Michael, R.P., Moore, K.L. & Cummings, R.D. Leukocyte trafficking mediated by selectin–carbohydrate interactions. *J. Biol. Chem.* **270**, 11025–11028 (1995).
49. Ye, Z. et al. ATP binding by monol-1/4NLRP12 is critical for its inhibitory function. *Mol. Cell. Biol.* **28**, 1841–1850 (2008).
50. Borghini, S. et al. Clinical presentation and pathogenesis of cold-induced autoinflammatory disease in a family with recurrence of an NLRP12 mutation. *Arthritis Rheum.* **63**, 830–839 (2011).
51. Landrum, M.J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014).
52. Arthur, J.C. Cutting edge: NLRP12 controls dendritic and myeloid cell migration to affect contact hypersensitivity. *J. Immunol.* **185**, 4515–4519 (2010).
53. Gaffin, S.L. An overview of IL-17 function and signaling. *Cytokine* **43**, 402–407 (2008).
54. Butcher, M.J., Gjurich, B.N., Phillips, T. & Galkina, E.V. The IL-17A/IL-17RA axis to therapy.
55. Butcher, M.J., Gjurich, B.N., Phillips, T. & Galkina, E.V. The IL-17A/IL-17RA axis to therapy.
56. Butcher, M.J., Gjurich, B.N., Phillips, T. & Galkina, E.V. The IL-17A/IL-17RA axis to therapy.
57. Butcher, M.J., Gjurich, B.N., Phillips, T. & Galkina, E.V. The IL-17A/IL-17RA axis to therapy.
58. Butcher, M.J., Gjurich, B.N., Phillips, T. & Galkina, E.V. The IL-17A/IL-17RA axis to therapy.
59. Butcher, M.J., Gjurich, B.N., Phillips, T. & Galkina, E.V. The IL-17A/IL-17RA axis to therapy.
60. Liu, J. et al. Piezo1 integration of vascular architecture with physiological force. *Nature* **515**, 279–282 (2014).
61. Ranade, S.S. et al. Piezo1, a mechanically activated ion channel, is required for vascular development in mice. *Proc. Natl. Acad. Sci. USA* **111**, 10347–10352 (2014).
62. Heyermet, R. et al. Piezo1-dependent stretch-activated channels are inhibited by Polycystin-2 in renal tubular epithelial cells. *EMBO Rep.* **14**, 1134–1148 (2013).
63. Miyamoto, T. et al. Functional role for Piezo1 in stretch-activated Ca<sup>2+</sup>-influx and ATP release in urethral cell cultures. *J. Biol. Chem.* **289**, 16565–16575 (2014).
64. Brohawn, S.G., Su, Z. & Mackinnon, R. Mechanosensitivity is mediated directly by the lipid membrane in TRAAK and TREK1 K<sup>+</sup> channels. *Proc. Natl. Acad. Sci. USA* **111**, 3614–3619 (2014).
65. Schröder, M.B., Nybo, M., Birgens, H. & Frederiksen, H. Hereditary xerocytosis and familial haemolysis due to mutation in the PIEZO1 gene: a simple diagnostic approach. *Int. J. Lab. Hematol.* **36**, e62–e65 (2014).
66. Yeo, N.C. et al. Shroom3 contributes to the maintenance of the glomerular filtration barrier integrity. *Genome Res.* **25**, 57–65 (2015).
67. Feluca, D. et al. Non-type 1 cytostatin caused by mutations in SLC7A9, encoding a subunit (bo,AT) of rBAT. *Nat. Genet.* **23**, 52–57 (1999).
68. Teslovich, T.M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
69. van den Berg, J.J. et al. Increased n-3 polysaturated fatty acid content of red blood cells from fish oil–fed rabbits increases in vitro lipid peroxidation, but decreases hemolysis. *Free Radic. Biol. Med.* **11**, 393–399 (1991).
70. Waldron, T. et al. C-N and its target Bmi1 are required for p190BCR/ABL leukemogenesis in mouse and human cells. *Leukemia* **26**, 644–653 (2012).
71. Schunkert, H. et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
72. Stahl, E.A. et al. Genome-wide association study identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508–514 (2010).
73. Levy, D. et al. Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**, 677–687 (2009).
74. Shameri, K. et al. A genome- and phenotype-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. *Hum. Genet.* **133**, 95–109 (2014).
75. Plagnol, V. et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genet.* **7**, e1002216 (2011).
76. Dichgans, M. et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke* **45**, 24–36 (2014).
ONLINE METHODS

Study samples. Our discovery sample consisted of exome chip data from 52,531 individuals, including 37,775 of European ancestry, 11,589 African Americans, and 3,167 Hispanic Americans sampled from 16 population-based cohorts participating in the CHARGE Consortium77: Age, Gene/Environment Susceptibility study (AGES), Atherosclerosis Risk in Communities (ARIC) Study, Cardiovascular Health Study (CHS), Family Heart Study (FamHS), Framingham Heart Study (FHS), Health ABC (HABC), Health2006/2008, the Mount Sinai Institute for Personalized Medicine BioMe Biobank Project (BioMe), the Jackson Heart Study (JHS), the Lothian Birth Cohorts 1921/1936 (LBC), the Multi-Ethnic Study of Atherosclerosis (MESA), the Rotterdam Study (RS), the Women's Health Initiative (WHI; African Americans only), and the Cardiovascular Risk in Young Finns Study (YFS). The replication sample consisted of 17,500 samples from the Women's Health Initiative (WHI; European Americans only) and 5,261 Han Chinese individuals from the Peking University–University of Michigan Study of Atherosclerosis (PUUMA).

Descriptions of each of the cohorts and the techniques used to measure the hematologic traits are provided in previous publications (Supplementary Note) and summarized in Supplementary Table 1. All participants provided written informed consent as approved by local human subjects committees.

Erythrocyte and leukocyte phenotypes. The hematology traits we studied included hemoglobin concentration (Hb), hematocrit (Hct), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), RBC count, RBC distribution width (RDW), total WBC count, and counts of the WBC subtypes neutrophils, monocytes, lymphocytes, basophils, and eosinophils, using the transformations defined in Supplementary Table 1. Traits were harmonized across cohorts for the same units of measurement, and within each cohort traits were transformed according to standard convention (Supplementary Table 1). We Winsorized values greater than 3 s.d. from the population mean for each trait in each cohort to reduce false positives caused by extreme outliers while still maintaining power to identify a potential signal with strong effect.

Genotyping and quality control. Genotypes were assayed using the Illumina HumanExome BeadChip in accordance with the manufacturer’s instructions. Genotype calls were assigned using GenomeStudio v2010.3. Samples were excluded if any of the following applied to them: a call rate <95%, ancestry outlier in a principal-components analysis, evidence of contamination, sex mismatch, or unexpected cryptic relatedness. SNPs were excluded with call rates <95% or if they deviated from Hardy–Weinberg equilibrium at P < 5 × 10⁻⁶. For the SNPs identified by the association analyses, the cluster plots were visually inspected.

Association analysis of single variants and implementation of gene-based tests. Variants were annotated using dbNSFP (v2.0)78-79. Phenotypes were first transformed (natural log transform, square root, or none at all, as delineated in Supplementary Table 1) and then Winsorized at 3 s.d. (mean and s.d. values were computed separately for each cohort and the threshold was computed as mean ± 3 s.d.; any individual with a value exceeding this threshold was replaced with this threshold). Age, sex, study (if needed), and principal components were included as covariates in the analyses. The R skatMeta (v1.4.2) package was used for all cohort-level analyses. Each study used either the skatCohort or the skatFamCohort function to create an R object that was then uploaded to a central server.

After performing quality control of the genotypes as described previously80, we analyzed 247,870 SNPs meeting quality control, using single-variant association tests and gene-based tests of aggregate variants. For single-variant association tests, a minor allele count of at least 40 was used for each trait. As a secondary analysis, we lowered this filter to a minor allele count of 10 or greater, to evaluate for any lower-frequency alleles with strong effects (Supplementary Fig. 8 and Supplementary Note). For gene-based testing, only coding variants putatively affecting protein structure (missense, stop-gain, stop-loss, and splice variants) that also had a frequency <5% in a given population (~200,000 SNPs) were included.

In parallel with the single-variant association tests, we conducted aggregate variant testing using two methods: the T5 test81 (MAF < 0.05) and the SKAT test82 (MAF < 0.05, Wu weights). The T5 test identifies genes where multiple samples have private or rare mutations leading to a strong effect in a single direction. The SKAT test allows for different variants to have effects in different directions. In both tests, only variants with a possible effect on amino acid sequence (missense, stop-gain, stop-loss, and splice variants) were included in the analysis.

Meta-analysis of single-variant and gene-based tests. Single-variant and gene-based association statistics were combined in a fixed-effects, inverse-variance-weighted meta-analysis performed in parallel at two different sites using the same skatMeta package. Analyses were first stratified by ancestry and then combined in a trans-ancestry analysis using the same methodology. Results for single-variant analyses were reported only when 40 or more minor allele counts were observed, and a Bonferroni correction for the number of tests was employed to determine significance. For gene-based tests, two different methods were employed. The first was the combined multivariate and collapsing (CMC) approach83, where the number of qualifying variants in each gene were added together for each individual separately and then used as the predictor in a linear regression model. To be included, a variant had to have an average allele frequency <5% across all cohorts and also change the amino acid sequence of an mRNA, either as a missense, stop-gain, stop-loss, frameshift, or splice-site variant. The second method was the SKAT method84 and used the same set of variants as the CMC/T5 approach. Only genes with a minor allele count >40 were analyzed, and a Bonferroni correction for the number of genes tested was employed to determine significance. The number of individuals with each of the hematologic traits under study differed, and consequently the number of markers reaching our minor allele count threshold of 40 varied by trait. We therefore applied trait-specific P-value thresholds, according to the number of variants available for the individuals with each trait (Supplementary Table 5).

Independent replication analysis. We conducted follow-up replication analysis in 18,018 independent European-American samples from the Women’s Health Initiative (WHI) and 5,261 Han Chinese individuals from the Shijingshan district of Beijing who participated in the Peking University–University of Michigan Study of Atherosclerosis (PUUMA) (Supplementary Note). Both studies were genotyped using an Illumina HumanExome BeadChip genotyping array and had erythrocyte and WBC traits available85. All new, significant (P value less than trait-specific Bonferroni threshold) variant associations from the discovery results were tested in the replication analysis. Gene-based test results that were significant in the discovery analyses were tested in the replication samples, with the exception of HFE and PIGM–DARC, as these loci have previously well-defined, known signals and were also seen in the single-variant analyses. In the case where an association was identified in the discovery analysis with an erythrocyte trait other than Hb or Hct, we analyzed the association with Hb and Hct in the replication analysis. Similarly, in the case where a leukocyte subtype association was found in the discovery analyses with an erythrocyte trait other than Hb or Hct, we analyzed the association with Hb and Hct in the replication analysis. We applied a Bonferroni correction to the number of replication tests we conducted for the single-variant analyses (P = 0.05/19 = 0.003) and for the gene-based tests (P = 0.05/10 = 0.005).

Expression quantitative trait locus analysis. We identified proxy SNPs in high LD (r² > 0.8) with associated index SNPs in HapMap 3 builds and the 1000 Genomes Project with SNAP83. SNP rsIDs were searched for primary SNPs and LD proxies against a collected database of expression SNP (eSNP) results (Supplementary Note). The collected eSNP results met criteria for statistical thresholds for association with gene transcript levels as described in the original papers.

Mouse experiments. Sldprδ−/− mice on a C57BL/6 background (stock 005799) were obtained from The Jackson Laboratory86. Mice were housed in a clean conventional facility that excluded specific mouse pathogens. All animal procedures were approved by the National Institute of Diabetes and Digestive and Kidney Diseases and were performed in accordance with the US National Institutes of Health guidelines. Because neutrophil counts are known to exhibit a high degree of variability within the same mouse and between mice, and by
sex84,85; we studied a total of 48 mice. The first 24 mice (6 S1pr4−/− females, 6 S1pr4−/− males, 6 S1pr4−/− females, and 6 S1pr4−/− males) were all littersmates (experiment 1 in Supplementary Table 14). In a second set of confirmatory experiments, 12 S1pr4−/− mice were compared to 12 C57BL/6 controls (Jackson Laboratory), again with equal proportions of males and females in each genotype group (experiment 2 in Supplementary Table 14). Mice were genotyped by multiplex PCR from tail snips using the set of primers and conditions previously described35. Mice were analyzed between 2 and 4 months after birth.

Total bone marrow cells were isolated from mice by flushing the femur and tibia from both legs two times with 1 ml of PBS. To obtain total leukocytes, the spleen was dissected and mechanically disaggregated. Single-cell suspensions were obtained using a 40-µm cell strainer. Blood samples were obtained by cardiac puncture. Erythrocytes were removed by ammonium chloride lysis. Absolute blood cell counts were determined by flow cytometry using CALTAG counting beads (Life Technologies), and the percentage of neutrophils of the total leukocyte pool was calculated and analyzed to account for any possible pipetting error. Neutrophils were analyzed by flow cytometry as previously described35. All antibodies were purchased from BD Bioscience and were used at a 1:50 dilution. Briefly, cells were diluted in 1% BSA-PBS and incubated with antibody to FcgR (553141 clone 2.4G2) followed by antibodies to mouse CD62L (553150 clone MEL-14), CD49d (CD11b (phycoerythrin (PE) conjugated; 553311 clone m1/70). Cells were also incubated with antibody to mouse CD62L (553310 clone MEL-14), CD49d (553156 clone R1-2), and CXCR4 (551967 clone 2B11/CXCR4) (all three antibodies were fluorescein conjugated). Cells were labeled for 30 min on ice and fixed in 1% paraformaldehyde in PBS and were then subjected to flow cytometry on a FACScalibur instrument (BD Bioscience). Data were analyzed using FlowJo software (Tree Star). Neutrophils were identified as Gr-1+CD11b+ cells, and monocyes were identified as Gr-1−CD11b+ cells.

Zebrafish experiments. Zebrafish ortholog s1pr4 was identified by sequence homology searches and gene synteny analysis, and MO design also incorporated information about gene structure and translational initiation sites (Gene-Tool). Two separate MOs were designed against s1pr4, which is a single-exon gene, in the ATG region to inhibit its mRNA translation (Supplementary Table 15). MOs were injected at multiple doses into one-cell-stage embryos of the mpx:gfp zebrafish line to find the optimal dose, 2 ng/embryo, and GFP-expressing cells were imaged under a spinning-disk confocal microscope (20x objective on an inverted Nikon Eclipse Ti microscope equipped with a Yokogawa spinning-disk confocal scan head and Andor iXon EM-CCD cameras) and counted at 2 d.p.f. Experiments were conducted in >10 embryos for each of the control and morphant constructs. Day 2 cutaneous injury was created 2 d after MO injection by nicking the tail fin, and the number of GFP+ cells at the site of the cutaneous wound was determined at 30 min and 1, 2, 3, 4, 5, 6, and 8 h after injury. Student’s t test was used to compare the neutrophil numbers at the cutaneous wound for the control and morphant constructs at each time point. Experiments were carried out in replicates of at least ten embryos by a technician, and analysis was checked by a postdoctoral fellow blinded to MO injection status.

77. 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 5 cohorts. Circ Cardiovasc Genet 2, 73–80 (2009).
78. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum. Mutat. 32, 894–899 (2011).
79. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: a database of human nonsynonymous SNVs and their functional predictions and annotations. Hum. Mutat. 34, E2393–E2402 (2013).
80. Grove, M.L. et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. PLoS One 8, e68095 (2013).
81. Li, B. & Leal, S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am. J. Hum. Genet. 83, 311–321 (2008).
82. Wu, M.C. et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. 89, 82–93 (2011).
83. Johnson, A.D. et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24, 2938–2939 (2008).
84. Bain, B.J. & England, J.M. Normal haematological values: sex difference in neutrophil count. BMJ 1, 306–309 (1975).
85. Bain, B.J. & England, J.M. Variations in leucocyte count during menstrual cycle. BMJ 2, 473–475 (1975).