Please cite this article in press as: Tajuddin et al., Large-Scale Exome-wide Association Analysis Identifies Loci for White Blood Cell Traits and Pleiotropy with..., The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.05.003

ARTICLE

Large-Scale Exome-wide Association Analysis Identifies Loci for White Blood Cell Traits and Pleiotropy with Immune-Mediated Diseases

Salman M. Tajuddin,[1,80] Ursula M. Schick,[2,3,80] John D. Eicher,[4] Nathalie Chami,[5,6] Ayush Giri,[7] Jennifer A. Brody,[8] W. David Hill,[9,10] Tim Kacprowski,[11,12] Jin Li,[13] Leo-Pekka Lyytikäinen,[14,15] Ani Manichaikul,[16] Evelin Mihailov,[17] Michelle L. O'Donoghue,[18] Nathan Pankratz,[19] Raha Pazoki,[20] Linda M. Polfus,[21] Albert Vernon Smith,[22,23] Claudia Schurmann,[2,3] Caterina Vacchi-Suzzi,[24] Dawn M. Waterworth,[25] Evangelos Evangelou,[26,27] Lisa R. Yanek,[28] Amber Burt,[29] Ming-Huei Chen,[4] Frank J.A. van Rooij,[20] James S. Floyd,[8] Andreas Greinacher,[30] Tamara B. Harris,[31] Heather M. Highland,[21,32] Leslie A. Lange,[33] Yongmei Liu,[34] Reedik Mägi,[17] Mike A. Nalls,[35] Rasika A. Mathias,[36] Deborah A. Nickerson,[37] Kjell Nikus,[38,39] John M. Starr,[9,40] Jean-Claude Tardif,[5,6] Ioanna Tzoulaki,[26,27] Digna R. Velez Edwards,[41] Lars Wallentin,[42] Traci M. Bartz,[43] Lewis C. Becker,[44] Joshua C. Denny,[45] Laura M. Raffield,[33] John D. Rioux,[5,6] Nele Friedrich,[12,46] Myriam Fornage,[47] He Gao,[26] Joel N. Hirschhorn,[48,49] David C.M. Liewald,[9,10] Stephen S. Rich,[16] Andre Uitterlinden,[20,50,51] Lisa Bastarache,[45] Diane M. Becker,[28] Eric Boerwinkle,[21,52] Simon de Denus,[6,53] Erwin P. Bottinger,[2]

(Author list continued on next page)

White blood cells play diverse roles in innate and adaptive immunity. Genetic association analyses of phenotypic variation in circulating white blood cell (WBC) counts from large samples of otherwise healthy individuals can provide insights into genes and biologic pathways involved in production, differentiation, or clearance of particular WBC lineages (myeloid, lymphoid) and also potentially inform the genetic basis of autoimmune, allergic, and blood diseases. We performed an exome array-based meta-analysis of total WBC and subtype counts (neutrophils, monocytes, lymphocytes, basophils, and eosinophils) in a multi-ancestry discovery and replication sample of ~157,622 individuals from 25 studies. We identified 16 common variants (8 of which were coding variants) associated with one or more WBC traits, the majority of which are pleiotropically associated with autoimmune diseases. Based on functional annotation, these loci included genes encoding surface markers of myeloid, lymphoid, or hematopoietic stem cell differentiation (*CD69*, *CD33*, *CD87*), transcription factors regulating lineage specification during hematopoiesis (*ASXL1*, *IRF8*, *IKZF1*, *JMJD1C*, *ETS2-PSMG1*), and molecules involved in neutrophil clearance/apoptosis (*C10orf54*, *LTA*), adhesion (*TNXB*), or centrosome and microtubule structure/function (*KIF9*, *TUBD1*). Together with recent reports of somatic *ASXL1* mutations among individuals with idiopathic cytopenias or clonal hematopoiesis of undetermined significance, the identification of a common regulatory 3′ UTR variant of *ASXL1* suggests that both germline and somatic *ASXL1* mutations contribute to lower blood counts in otherwise asymptomatic individuals. These association results shed light on genetic mechanisms that regulate circulating WBC counts and suggest a prominent shared genetic architecture with inflammatory and autoimmune diseases.

Introduction

White blood cells (WBCs) are major constituents of the blood and lymphatic system. They are classified into two lineages: myeloid (neutrophils, basophils, eosinophils, and monocytes) and lymphoid (lymphocytes). Lineage commitment of hematopoietic stem cells involves precise transcriptional and epigenetic regulation, creating the

[1]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Baltimore, MD 21224, USA; [2]The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [3]The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [4]Population Sciences Branch, National Heart Lung and Blood Institute, The Framingham Heart Study, Framingham, MA 01702, USA; [5]Department of Medicine, Université de Montréal, Montréal, QC H3T 1J4, Canada; [6]Montreal Heart Institute, Montréal, QC H1T 1C8, Canada; [7]Division of Epidemiology, Institute for Medicine and Public Health, Vanderbilt University, Nashville, TN 37235, USA; [8]Department of Medicine, University of Washington, Seattle, WA 98101, USA; [9]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK; [10]Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK; [11]Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald and Ernst-Mortiz-Arndt University Greifswald, Greifswald 17475, Germany; [12]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, Germany; [13]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Palo Alto, CA 94305, USA; [14]Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland; [15]Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere 33014, Finland; [16]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA; [17]Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; [18]TIMI Study Group, Cardiovascular Division, Brigham and Women's Hospital, Boston, MA 02115, USA; [19]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55454, USA; [20]Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3000, the Netherlands; [21]Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [22]Icelandic Heart Association, 201 Kopavogur, Iceland; [23]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; [24]Department of Family,

(Affiliations continued on next page)

© 2016 American Society of Human Genetics.

Caroline Hayward,[54] Albert Hofman,[20,55] Georg Homuth,[11] Ethan Lange,[56] Lenore J. Launer,[31] Terho Lehtimäki,[14,15] Yingchang Lu,[2,3] Andres Metspalu,[17] Chris J. O'Donnell,[57,58] Rakale C. Quarells,[59] Melissa Richard,[47] Eric S. Torstenson,[7] Kent D. Taylor,[60,61] Anne-Claire Vergnaud,[26] Alan B. Zonderman,[1] David R. Crosslin,[62] Ian J. Deary,[9,10] Marcus Dörr,[12,63] Paul Elliott,[26] Michele K. Evans,[1] Vilmundur Gudnason,[22,23] Mika Kähönen,[64,65] Bruce M. Psaty,[66,67] Jerome I. Rotter,[60,61] Andrew J. Slater,[68,69] Abbas Dehghan,[20] Harvey D. White,[70] Santhi K. Ganesh,[71] Ruth J.F. Loos,[2,3,72] Tõnu Esko,[17,48] Nauder Faraday,[73] James G. Wilson,[74] Mary Cushman,[75] Andrew D. Johnson,[4] Todd L. Edwards,[76] Neil A. Zakai,[75] Guillaume Lettre,[5,6,80] Alex P. Reiner,[77,78,80,*] and Paul L. Auer[79,80,*]

specific bone marrow microenvironment to produce each distinct mature blood cell type.[1] Mature WBCs play diverse, choreographed roles in innate and adaptive immunity including detection, neutralization, and elimination of invading pathogens, response to tissue injury, and wound healing. In addition, WBCs are associated with the development of chronic inflammatory, allergic, and autoimmune diseases.[2] Therefore, total and differential WBC counts are important clinical measures of susceptibility to infection and used to monitor disease activity and tolerability to therapeutic regimens for oncologic and rheumatologic diseases.

Total and differential WBC counts are complex, polygenic traits with estimated heritability of 50%–60%.[3] Previous genome-wide association studies (GWASs) have characterized common and lower frequency variation contributing to WBC counts in European, African, and Asian ancestry populations (N.P., U.M.S., J.B.-J., and M.-H.C., unpublished data).[3–12] More than 30 distinct genetic loci have been discovered; in some instances, these genetic studies have provided important biologic insights into the development, maturation, or regulation of WBC types. Nonetheless, these studies have explained only a small proportion (<10%) of the estimated heritability of

Population and Preventive Medicine, Stony Brook University, Stony Brook, NY 11794, USA; [25]Genetics, Target Sciences, GlaxoSmithKline, King of Prussia, PA 19406, USA; [26]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, UK; [27]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina 45110, Greece; [28]Department of Medicine, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [29]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA 98195, USA; [30]Institute for Immunology and Transfusion Medicine, University Medicine Greifswald, Greifswald 17475, Germany; [31]Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, Intramural Research Program, NIH, Bethesda, MD 20892, USA; [32]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA; [33]Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA; [34]Center for Human Genetics, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; [35]Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD 20892, USA; [36]Department of Medicine, Divisions of Allergy and Clinical Immunology and General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [37]Department of Genome Sciences, School of Medicine, University of Washington, Seattle, WA 98105, USA; [38]Department of Cardiology, Heart Center, Tampere University Hospital, Tampere 33521, Finland; [39]University of Tampere School of Medicine, Tampere 33014, Finland; [40]Alzheimer Scotland Dementia Research Centre, Edinburgh EH8 9JZ, UK; [41]Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN 37203, USA; [42]Department of Medical Sciences, Cardiology, and Uppsala Clinical Research Center, Uppsala University, 751 85 Uppsala, Sweden; [43]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [44]Department of Medicine, Divisions of Cardiology and General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [45]Department of Biomedical Informatics, School of Medicine, Vanderbilt University, Nashville, TN 37203, USA; [46]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald 13347, Germany; [47]Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [48]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA; [49]Department of Endocrinology, Boston Children's Hospital, Boston, MA 02115, USA; [50]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam 3000, the Netherlands; [51]Netherlands Consortium for Healthy Ageing (NCHA), Rotterdam 3015, the Netherlands; [52]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [53]Faculty of Pharmacy, Université de Montréal, Montréal, QC H3T 1J4, Canada; [54]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK; [55]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA 02115, USA; [56]Departments of Genetics and Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [57]National Heart, Lung, and Blood Institute, The Framingham Heart Study, Framingham, MA 01702, USA; [58]Cardiology Section and Center for Population Genomics, Boston Veteran's Administration (VA) Healthcare, Boston, MA 02118, USA; [59]Morehouse School of Medicine, Social Epidemiology Research Center, Cardiovascular Research Institute, Atlanta, GA 30310, USA; [60]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute, Torrance, CA 90502, USA; [61]Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [62]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA 98195, USA; [63]Department of Cardiology, University Medicine Greifswald, Greifswald 17475, Germany; [64]Department of Clinical Physiology, Tampere University Hospital, Tampere 33521, Finland; [65]Department of Clinical Physiology, University of Tampere School of Medicine, Tampere 33014, Finland; [66]Cardiovascular Health Research Unit, Departments of Epidemiology, Health Services, and Medicine, University of Washington, Seattle, WA 98101, USA; [67]Group Health Research Institute, Group Health Cooperative, Seattle, WA 98101, USA; [68]OmicSoft Corporation, Cary, NC 27513, USA; [69]Genetics, Target Sciences, GlaxoSmithKline, Research Triangle Park, NC 27709, USA; [70]Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland 1142, New Zealand; [71]Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI 48108, USA; [72]The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [73]Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [74]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA; [75]Division of Hematology Oncology, Department of Medicine, The University of Vermont, Colchester, VT 05446, USA; [76]Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN 37203, USA; [77]Department of Epidemiology, University of Washington, Seattle, WA 98195, USA; [78]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [79]Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI 53205, USA
[80]These authors contributed equally to this work
*Correspondence: apreiner@u.washington.edu (A.P.R.), pauer@uwm.edu (P.L.A.)
http://dx.doi.org/10.1016/j.ajhg.2016.05.003.

Table 1. Sample Sizes for Exome-wide Association Analyses of White Blood Cell Traits

Population	Total WBC	Neutrophils	Monocytes	Lymphocytes	Basophils	Eosinophils
Discovery						
European ancestry	108,596	60,851	44,325	47,105	44,138	32,517
African ancestry	23,250	10,119	9,790	9,808	9,509	8,282
Hispanic American	5,536	4,825	3,452	3,450	3,453	3,450
East Asian	968	965	–	–	–	–
South Asian	464	463	–	–	–	–
Replication						
European ancestry	18,808	17,066	17,066	17,109	16,189	15,327
Total	157,622	94,289	74,633	77,472	73,289	59,576

WBC traits in European ancestry populations[6] and less than 25% in African ancestry (AA) populations (in AA, a substantial proportion of the variation in WBC counts is attributed to a single variant—rs2814778—in *DARC* [Duffy Antigen Receptor for Chemokines (MIM: 613665)]).[3,13] In an effort to augment the discoveries from GWASs and to identify additional functional loci contributing to variation in WBC counts, we performed exome array-based meta-analysis of total and differential counts in a multi-ancestry samples from 25 studies.

Material and Methods

Study Subjects

The Blood-Cell Consortium (BCX) is an international collaboration with the goal of identifying common and rare variants associated with blood cell traits through exome genotyping arrays (Table S1). The consortium, which is comprised of multi-ancestry cohorts including European ancestry (EA), African ancestry (AA), Hispanic ancestry (HA), East Asian ancestry (EAS), and South Asian ancestry (SA), is divided into three main working groups: red blood cell (RBC), platelet, and WBC. For exome-wide association analysis of WBC traits, the discovery and replication phases included a total of 157,622 participants from 25 cohorts (Tables 1, S2, and S3). The discovery sample consisted of up to 138,814 individuals from 21 studies. The replication sample included 18,808 independent individuals from 4 additional studies. The division of discovery and replication samples was dictated by timing; we collected all available studies for initial discovery and then identified others who could participate only at a later point in time and hence were used for replication. A summary of descriptive statistics for total WBC, neutrophils, monocytes, lymphocytes, basophils, and eosinophils is shown in Table S4. All participants provided informed consent and the study was approved by the Institutional Review Board of each participating study.

Genotyping and Quality Control

Each participating study used one of the following exome content genotyping arrays: Illumina ExomeChip v.1.0, Illumina ExomeChip v.1.1_A, Illumina ExomeChip-12 v.1.1, Affymetrix Axiom Biobank Plus GSKBB1, or Illumina HumanOmniExpressExome Chip. Genotypes were called either using a combination of the Illumina GenomeStudio and zCall software or using the Exomechip joint calling plan developed by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium[14] (Table S1). Standard quality-control criteria were applied by each study. Exclusion criteria included sample call rates of less than 98%, excess heterozygosity rates, Hardy-Weinberg equilibrium p values $< 1 \times 10^{-6}$, and sex mismatch. Additionally, ancestry was confirmed through principal components or multi-dimensional scaling analyses using linkage disequilibrium (LD) pruned markers ($r^2 < 0.2$) with minor allele frequency greater than 1%. Scatterplots anchored using the 1000 Genomes Project populations were visually inspected and ancestry outliers were excluded. Insertion and deletion variants and variants mapping to the Y chromosome, the pseudo-autosomal region, or mitochondrial sequence were removed, leaving only those on the autosomal and X chromosomes. All remaining variants (including monomorphic variants) were aligned to the forward strand and alleles were checked to ensure that the correct reference allele was specified. After all quality-control procedures, each study generated an indexed variant call file (VCF) for subsequent analyses. The VCF files were checked for allele alignment with the checkVCF package.

We performed study-specific quality control on each trait association result using the EasyQC protocol.[15] Variant allele frequencies from each study were plotted against ethnicity-specific reference population allele frequencies to identify allele frequency deviations and the presence of flipped alleles. In order to assess proper trait transformation in each cohort, a scatterplot of the median standard error versus study-specific sample size was visually inspected for deviations.

Statistical Analysis

To assess the association between WBC-related traits and Exomechip variants, white blood cell and differential counts (total WBC, neutrophils, monocytes, lymphocytes, eosinophils, and basophils) were obtained from complete blood cell count. Each of the WBC-related traits was $\log_{10}$ transformed to normalize the distribution of the traits. In each participating study, residuals for each WBC trait were calculated from linear regression models adjusted for age, age-squared, sex, study center (where applicable), and principal components. Residuals from this model were then transformed using the rank-based inverse normal transformation to control type I error.[16] Autosomal and X chromosome variants were then tested for association with each WBC trait using either Rvtests or RAREMETALWORKER software packages. Both packages generate single variant association score summary statistics, variance-covariance matrices containing LD relationships between

variants within a 1 MB window, and variant-specific parameters including minor allele frequency, chromosome position, strand, genotype call rate, and Hardy-Weinberg equilibrium p values.

Discovery Association Meta-analysis

For each WBC trait, we performed three distinct discovery meta-analyses: in EA only, AA only, and combined across all five ancestry groups. Ancestry-stratified (EA and AA) and combined all ancestry (EA, AA, HA, EAS, and SA) meta-analyses of single variant association results were carried out using the Cochran-Mantel-Haenszel approach implemented in RareMETALS.[17] We included variants in the meta-analysis if the genotype call rate was $\geq$95%. For palindromic variants (i.e., A/T and C/G variants), we compared allele frequencies taken across the entire consortium in order to detect flipped alleles. We kept variants with an allele frequency difference <0.3 or <0.6 for ancestry-specific (EA, AA) or combined all ancestry analyses, respectively.[15] Using single-variant score statistics and variance-covariance matrices of LD estimates, we performed two types of gene-based tests across the contributing studies: (1) a burden test that assumes all qualifying rare variants in a gene are associated with a trait with the same direction of effect (variable threshold test), and (2) the sequence kernel association test (SKAT) that accounts for rare variants in a gene having opposing direction of effects.[17] For all gene-based tests performed, we considered single-nucleotide variants (SNVs) with an allele frequency of $\leq$1% and annotated as missense, nonsense, and splice site variants; the latter two categories include loss-of-function variants. Similar to the single-variant analyses, results were generated for EA, AA, and for the combined all ancestry samples. For the discovery single variant and gene-based association analyses, the statistical significance threshold was set as p value $< 2 \times 10^{-7}$ and $< 3 \times 10^{-6}$, respectively.

Conditional Analysis

To identify multiple independent associations within a region, using the RareMETALS software we performed stepwise conditional analyses adjusting for the most significant single variant in a 1 MB window, across the entire Exomechip array. This step was repeated until there was no new association signals identified in each region, defined as a p value $< 2 \times 10^{-7}$. Further, to assess whether SNVs identified by the present study were independent of any previously reported WBC-associated variants, we conditioned our regression models on known GWAS sentinel variants or their proxies (LD $r^2 \geq 0.80$). For regions of the genome where there is extended LD structure spanning more than 1 MB, we performed a stepwise conditional analysis in GCTA software[18] conditioning on the most significant variant in the region first (or the GWAS sentinel variant or LD proxy).

Replication Meta-analysis

We sought replication of association results using four independent European ancestry cohorts (Tables 1 and S3). The single-variant association results from each replication cohort were combined using the Cochran-Mantel-Haenszel method in RareMETALS. Contributing replication cohorts adhered to the quality control and association analysis procedures described previously for the discovery analysis. Replication of association findings were considered significant if the variants demonstrated the same direction of effect as the discovery association meta-analyses with a replication p value < 0.05. A meta-analysis of discovery and replication results was performed using an inverse-variance weighting method as implemented in METAL.[19] We also performed replication of gene-based associations in independent ~2,900 EA samples.

Phenome-wide Association Study Analysis

In 29,722 EA samples from the BioVU study,[20] we performed phenome-wide association study (PheWAS) analysis[21] to assess the association between our WBC-related loci and 1,502 International Classification of Disease, Ninth Revision (ICD-9) code curated clinical phenotypes.[21] Variants were included in the analysis if there were ten cases with at least one copy of the minor allele. Associations between SNVs and phenotypes were assessed using a logistic regression model adjusted for sex and five principal components. Empirical significance was estimated by permutation test. The permutation test was performed by assigning each vector of clinical phenotypes to a random subject 50,000 times, and then scanning all SNV-phenotype combinations with association tests. We then created a ranked distribution of the maximum test statistics over all SNV-phenotype combinations in each of the 50,000 permutations. The 95th percentile of the distribution of maximum test statistics across the 1,502 clinical phenotypes and 95 SNVs equates to a threshold that controls the family-wise error rate at 0.05. This threshold accounts for multiple testing across SNVs and phenotypes. Our observed test statistics greater than this 95th percentile were considered statistically significant.

To further assess pleiotropy between WBC-associated variants and inflammatory diseases, we performed lookups in published GWASs of various autoimmune diseases (celiac disease [MIM: 212750], inflammatory bowel disease [IBD; MIM: 266600], multiple sclerosis [MS; MIM: 126200], primary biliary cirrhosis [PBC; MIM: 109720], psoriasis [MIM: 177900], rheumatoid arthritis [RA; MIM: 180300], systemic lupus erythematosus [SLE; MIM: 152700], type 1 diabetes mellitus [T1D; MIM: 222100]) and coronary artery disease (MIM: 608901).[22–30] We supplemented the full GWAS summary statistics lookups with the GRASP database[31] to include other immunologically relevant clinical phenotypes and quantitative traits. Similarly, to assess whether the WBC variants were associated with other blood cell traits, we obtained effect sizes and p values for these variants from RBC- and platelet-related traits exome array analyses within the BCX consortium.[32,33]

Functional Annotation of Variants

To assess the functional consequences of coding and non-coding variants associated with WBC traits, we utilized a variety of existing variant annotation resources. Using a curated collection of more than 100 separate expression quantitative trait loci (eQTL) datasets, we queried whether our list of WBC-trait loci were also associated with transcript expression in blood-cell-specific eQTL datasets. A general overview of a subset of >50 eQTL studies has been published,[34] with specific citations for the blood-cell-specific eQTL datasets shown in Table S5. Additional in silico functional annotations were performed with ANNOVAR.[35] The deleteriousness of each variant was estimated with the Combined Annotation-Dependent Depletion (CADD) score where each variant is assigned a scaled C score; a score of greater than 10 is suggested to indicate deleteriousness.[36]

Results

We conducted an exome-wide association analyses of total WBC and differential counts (neutrophils, monocytes,

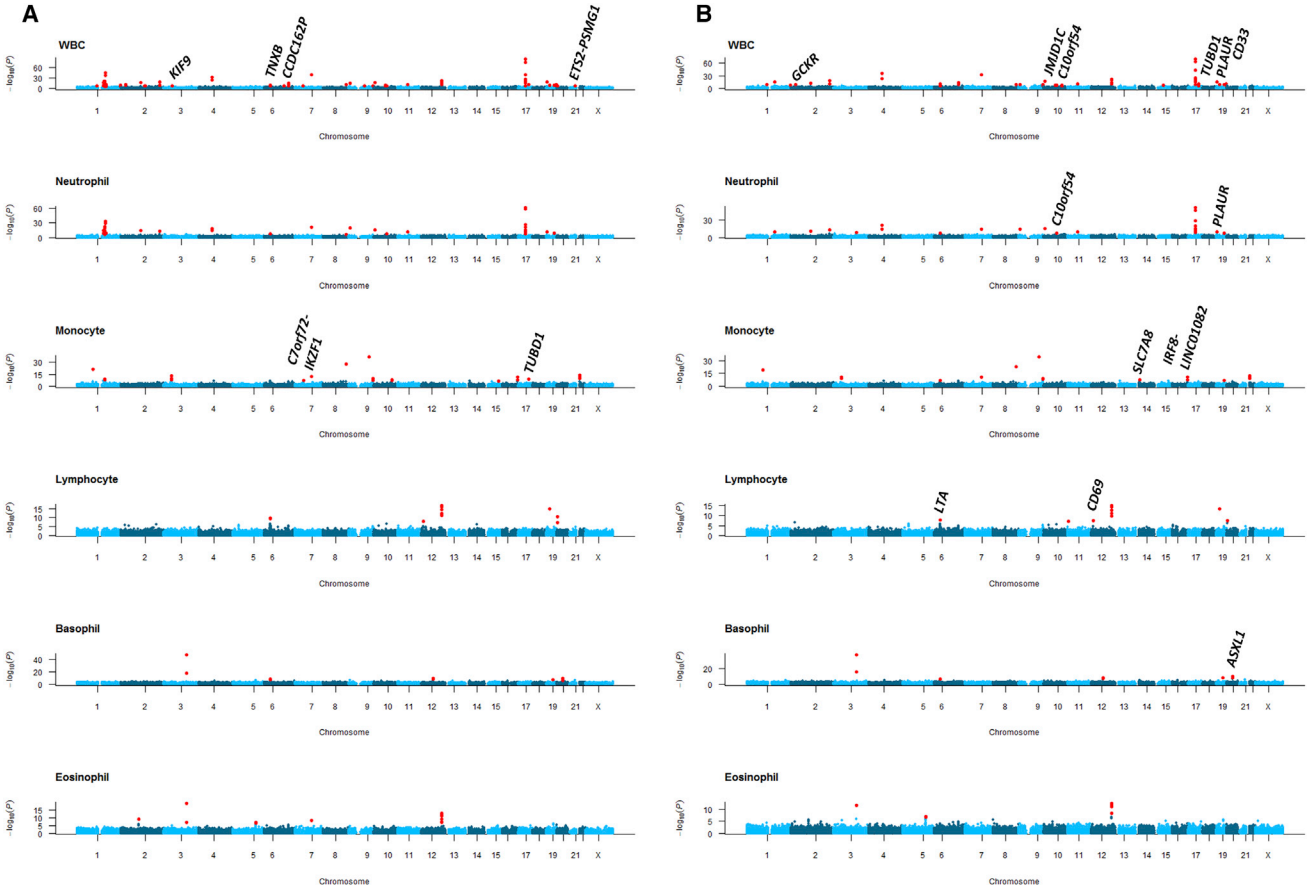


Figure 1. Manhattan Plots of p Values of White Blood Cell Traits
(A) Discovery association results in the combined all ancestries sample.
(B) Discovery association results in the European ancestry samples.
The combined all ancestry sample include European, African, Hispanic, East Asian, and South Asian ancestries. Genetic variants that passed the array-wide significance threshold (p value $< 2.0 \times 10^{-7}$) are highlighted in red. Discovery genetic loci that replicated in independent samples are shown.

lymphocytes, basophils, and eosinophils) in a discovery sample of ~138,814 individuals of European, African, Hispanic, East Asian, and South Asian ancestries across 21 cohorts (Tables 1 and S3). Quantile-quantile plots with genomic inflation factors and their respective Manhattan plots for each discovery meta-analysis are presented in Figures 1, S1, and S2. The discovery effort yielded 144 array-wide significant SNV associations (p value $< 2.0 \times 10^{-7}$) (Table S6). After stepwise conditional analyses, we refined this list to 28 independent SNV associations with WBC counts that were not previously reported (Table S7). Of these 28 variant associations, 16 were replicated (p value < 0.05 and consistent direction of effect) in 17,897 independent EA individuals (Figure 1, Table 2). Fourteen of the replicated loci are located in genomic regions not previously associated with WBC traits. The remaining two loci (*TNXB* rs185819 and *IRF8* rs11642873) represent secondary, independent signals located within a 1 MB window of a previously reported WBC locus. Of the 16 replicated loci, 10 were significantly associated with total WBC count, 2 with neutrophil count, 4 with monocyte count, 2 with lymphocyte count, and 1 with basophil count. As described further below, several loci were associated with more than one WBC trait (Table 2); the WBC-subtype-specific association results for each of the 16 replicated variants are shown in Table S8. For each locus, the allele frequencies stratified by ancestry are shown in Table S9. The full summary Exomechip association results for all traits are publicly available online (see Web Resources).

Total WBC

We found missense variants in a number of genes that were associated with total WBC. In *GCKR* (MIM: 600842), rs126032 (p.Leu446Pro [c.1337T>C]) was associated with lower total WBC in the EA meta-analysis (p value $= 8.13 \times 10^{-13}$). This variant was also nominally associated with lower neutrophil, lymphocyte, and basophil counts in EAs, consistent with its association with total WBC. The rs126032 variant was also associated with lower total WBC in AAs (p value $= 0.014$). In *KIF9* (MIM: 607910), rs2276853 (p.Arg573Trp [c.1717C>T]) was associated with increased total WBC in the multi-ancestry meta-analysis (p value $= 3.29 \times 10^{-9}$). The signal was largely driven

Please cite this article in press as: Tajuddin et al., Large-Scale Exome-wide Association Analysis Identifies Loci for White Blood Cell Traits and Pleiotropy with..., The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.05.003

Table 2. Variants Associated with White Blood Cell Traits

Trait (Population)	dbSNPID	Chr	Pos	Alt/ Ref	EAF	Gene	Annotation	AA Substitution	Discovery N	Discovery Beta (SE)	Discovery p	Replication N	Replication EAF	Replication Beta (SE)	Replication p	Combined Meta-analysis N	Combined Meta-analysis EAF	Combined Meta-analysis Beta (SE)	Combined Meta-analysis p	p_{het}
WBC (EA)	rs1260326	2	27,730,940	C/T	0.58	*GCKR*	missense, splice site	p.Leu446Pro	108,596	−0.030 (0.005)	4.01×10^{-10}	17,897	0.6	−0.044 (0.012)	2.66×10^{-4}	126,493	0.58	−0.032 (0.004)	8.13×10^{-13}	0.28
WBC (All)	rs2276853	3	47,282,303	A/G	0.58	*KIF9*	missense	p.Arg573Trp	132,764	0.023 (0.004)	3.65×10^{-8}	17,897	0.6	0.025 (0.012)	3.00×10^{-2}	150,661	0.58	0.023 (0.004)	3.29×10^{-9}	0.86
WBC (All)	rs185819[a]	6	32,050,067	C/T	0.51	*TNXB*	missense	p.His1161Arg	132,764	0.031 (0.005)	4.02×10^{-10}	17,897	0.47	0.034 (0.015)	2.24×10^{-2}	150,661	0.51	0.031 (0.005)	2.85×10^{-11}	0.83
WBC (All)	rs9374080	6	109,616,420	C/T	0.43	*CCDC162P*	intronic regulatory		132,764	0.023 (0.004)	4.01×10^{-8}	17,897	0.46	0.025 (0.011)	2.55×10^{-2}	150,661	0.43	0.023 (0.004)	3.15×10^{-9}	0.84
WBC (EA)	rs3747869	10	73,520,632	C/A	0.9	*C10orf54 (DD1a)*	missense	p.Asp187Glu	108,596	0.040 (0.007)	4.26×10^{-8}	17,897	0.9	0.083 (0.018)	6.40×10^{-6}	126,493	0.9	0.046 (0.007)	1.42×10^{-11}	0.03
WBC (EA)	rs1935	10	64,927,823	G/C	0.49	*JMJD1C*	missense	p.Glu2353Asp	108,596	−0.026 (0.005)	2.46×10^{-8}	17,897	0.46	−0.027 (0.012)	2.06×10^{-2}	126,493	0.49	−0.026 (0.004)	1.57×10^{-9}	0.93
WBC (EA)	rs1292053	17	57,963,537	G/A	0.45	*TUBD1*	missense	p.Met76Thr	108,596	−0.03 (0.004)	1.28×10^{-11}	17,897	0.44	−0.027 (0.011)	1.51×10^{-2}	126,493	0.45	−0.030 (0.004)	6.55×10^{-13}	0.78
WBC (EA)	rs4760	19	44,153,100	G/A	0.16	*CD87 (PLAUR)*	missense	p.Leu272Pro	85,685	−0.043 (0.007)	2.51×10^{-10}	17,897	0.15	−0.052 (0.015)	7.13×10^{-4}	103,582	0.16	−0.044 (0.006)	8.34×10^{-13}	0.6
WBC (EA)	rs3865444	19	51,727,962	A/C	0.31	*CD33*	upstream		86,936	−0.037 (0.005)	3.51×10^{-12}	17,897	0.32	−0.033 (0.012)	5.14×10^{-3}	104,833	0.31	−0.036 (0.005)	6.81×10^{-14}	0.77
WBC (All)	rs2836878	21	40,465,534	A/G	0.26	*ETS2-PSMG1*	intergenic		132,764	−0.025 (0.005)	8.36×10^{-8}	17,897	0.26	−0.026 (0.012)	3.44×10^{-2}	150,661	0.26	−0.025 (0.004)	8.41×10^{-9}	0.89
NEU (EA)	rs3747869	10	73,520,632	C/A	0.9	*C10orf54 (DD1a)*	missense	p.Asp187Glu	60,851	0.053 (0.010)	2.11×10^{-8}	16,669	0.9	0.073 (0.019)	1.17×10^{-4}	77,520	0.9	0.057 (0.009)	1.65×10^{-11}	0.34
NEU (EA)	rs4760	19	44,153,100	G/A	0.16	*CD87 (PLAUR)*	missense	p.Leu272Pro	56,112	−0.047 (0.008)	1.54×10^{-8}	16,669	0.15	−0.044 (0.016)	5.55×10^{-3}	72,781	0.16	−0.046 (0.007)	3.01×10^{-10}	0.87
MON (All)	rs4917014	7	50,305,863	G/T	0.28	*C7orf72-IKZF1*	intergenic		57,183	−0.038 (0.007)	1.97×10^{-8}	16,669	0.32	−0.048 (0.012)	8.92×10^{-5}	73,852	0.29	−0.040 (0.006)	9.75×10^{-12}	0.48
MON (EA)	rs11625112	14	23,596,740	G/A	0.46	*SLC7A8*	intronic		44,325	−0.038 (0.007)	3.82×10^{-8}	16,669	0.45	−0.031 (0.012)	7.04×10^{-3}	60,994	0.46	−0.036 (0.006)	1.03×10^{-9}	0.62
MON (EA)	rs11642873[a]	16	85,991,705	C/A	0.2	*IRF8-LINC01082*	intergenic		44,325	0.057 (0.008)	1.41×10^{-11}	16,669	0.2	0.113 (0.014)	6.17×10^{-15}	60,994	0.2	0.072 (0.007)	1.40×10^{-22}	0.001
MON (All)	rs1292053	17	57,963,537	G/A	0.45	*TUBD1*	missense	p.Met76Thr	57,183	−0.036 (0.006)	2.55×10^{-9}	16,669	0.44	−0.040 (0.012)	6.08×10^{-4}	73,852	0.45	−0.037 (0.005)	6.53×10^{-12}	0.76
LYM (EA)	rs2229094	6	31,540,556	C/T	0.26	*LTA*	missense	p.Cys13Arg	47,105	0.046 (0.008)	1.89×10^{-8}	16,711	0.25	0.078 (0.018)	8.54×10^{-6}	63,816	0.26	0.051 (0.007)	3.14×10^{-12}	0.09

(*Continued on next page*)

Table 2. *Continued*

Trait (Population)	dbSNPID	Chr	Pos	Alt/ Ref	EAF	Gene	Annotation	AA Substitution	Discovery			Replication					Combined Meta-analysis				
									N	Beta (SE)	p	N	EAF	Beta (SE)	p		N	EAF	Beta (SE)	p	p_{het}
LYM (EA)	rs4763879	12	9,910,164	A/G	0.36	*CD69*	intronic regulatory		47,105	−0.038 (0.007)	3.08×10^{-8}	16,711	0.36	−0.038 (0.012)	1.36×10^{-3}		63,816	0.36	−0.038 (0.006)	1.59×10^{-10}	0.99
BAS (EA)	rs2295764	20	31,025,163	G/A	0.36	*ASXL1*	3′ UTR		44,138	−0.042 (0.007)	3.28×10^{-9}	15,770	0.36	−0.031 (0.012)	9.78×10^{-3}		59,908	0.36	−0.039 (0.006)	1.46×10^{-10}	0.43

Abbreviations are as follows: Chr, chromosome; Pos, basepair position; Alt, effect allele; Ref, reference allele; EAF, effect allele frequency; AA, amino acid; p_{het}, p value for heterogeneity; EA, European ancestry; All, combined European, African, Hispanic American, East Asian, and South Asian ancestries; WBC, white blood cell; NEU, neutrophil; MON, monocyte; LYM, lymphocyte; BAS, basophil.
[a]Secondary signal identified through conditional analysis.

by the association in EAs (p value = 1.39×10^{-6}) and was apparent for both neutrophil and lymphocyte counts in EAs and in multi-ancestry meta-analyses. In *TNXB* (MIM: 600985), rs185819 (p.His1161Arg [c.3428A>G]) was associated with increased total WBCs in the multi-ancestry meta-analysis (p value = 2.85×10^{-11}). The association was consistently significant across EA and AA populations and for all WBC sub-types. The effect allele frequency was comparable between EAs and AAs but varied in the other ancestry groups. In *C10orf54* (MIM: 615608), rs3747869 (p.Asp187Glu [c.561T>G]) was associated with increased total WBC in the EA meta-analysis (p value = 1.42×10^{-11}). Although rs3747869 was also associated with neutrophil, monocyte, and eosinophil counts, the signal was not consistent across ancestry groups. The effect allele frequencies were markedly different between EA, AA, HA, SA, and EAS ancestry groups. In *JMJD1C* (MIM: 604503), rs1935 (p.Glu2353Asp [c.7059G>C]) was associated with lower total WBC (p value = 1.57×10^{-9}) in the EA meta-analysis. Although the rs1935 variant was not consistently associated with total WBC across all the major ethnic groups, it was significant in the HAs (p value = 5.58×10^{-3}). Significantly low neutrophil, lymphocyte, and eosinophil counts were also observed for rs1935. In *TUBD1* (MIM: 607344), rs1292053 (p.Met76Thr [c.227T>C]) was associated with lower total WBC in the EA meta-analysis (p value = 6.55×10^{-13}). This association was similar in EAs and AAs and for neutrophil, monocyte, and lymphocyte counts. Finally, in *PLAUR* (MIM: 173391) the rs4760 (p.Leu272Pro [c.815T>C]) variant was associated with lower total WBC (p value = 8.34×10^{-13}) in the EA meta-analysis. The effect allele frequencies were highly discrepant across ancestries, perhaps explaining why the association was observed only in EAs. The rs4760 association with total WBC was almost entirely due to its strong association with neutrophil counts.

Outside of coding regions, an intronic variant (rs9374080) in *CCDC162P* was associated with increased total WBC in the multi-ancestry meta-analysis (p value = 3.15×10^{-9}). The association was consistent across EAs and AAs and was observed for neutrophil and monocyte counts and was especially strong for basophil counts. The rs3865444 variant, just upstream of *CD33* (MIM: 159590), was associated with lower total WBC in the EA meta-analysis (p value = 6.81×10^{-14}). The allele frequencies were highly discrepant across ancestry groups and rs3865444 was not significantly associated with total WBC outside of the EAs. However, the association was consistent across neutrophil, monocyte, and eosinophil counts. Finally, an intergenic variant (rs2836878) near *ETS2* (MIM: 164740) and *PSMG1* (MIM: 605296) was associated with lower total WBC in the multi-ancestry meta-analysis (p value = 8.41×10^{-9}). The association was driven by the EA signal, and the variant had different allele frequencies across ancestry groups. The association with total WBC was consistent across neutrophil, monocyte, and basophil counts.

We identified a rare, missense variant in *OR4C6* (rs144349650, p.Leu112Val [c.334C>G], EAF = 0.00042) that was significantly associated with lower total WBC in the EA discovery analysis (p value = 1.87×10^{-11}; Table S7). The allele frequency was rare in all ancestry groups and did not replicate in additional samples of >17,000 EAs, perhaps due to low statistical power. Likewise, we identified a burden of rare, missense variants in *TAF3* (MIM: 606576) that was significantly associated with increased total WBC in the EA discovery set (p_{VT} = 1.58×10^{-6}; Table S10). However, the signal did not replicate in an additional independent 2,898 samples.

Neutrophil Count

In addition to the associations with total WBC, we identified two missense variants that were associated with neutrophil count at exome-wide significance levels. The effect estimate of the rs3747869 variant in *C10orf54* for total WBC appeared to be a combination of effects from neutrophil, monocyte, and eosinophil counts, though the effect was strongest for neutrophils, largely explaining the overall association with total WBC. The association between rs4760 in *PLAUR* and total WBC also appeared to be explained by the association with neutrophil counts.

The association between the rare, missense rs144349650 variant in *OR4C6* was observed for neutrophil counts as well as total WBC in the EA and multi-ancestry discovery sets. In gene-based test, *OR4C6* was associated with neutrophil count ($p_{SKAT} = 2.56 \times 10^{-8}$; Table S10). Likewise, a burden of rare, missense variants in *ZNF439* was associated for neutrophil counts in the AA set ($p_{VT} = 9.57 \times 10^{-7}$; Table S10). Neither the *ZNF439* nor the *OR4C6* gene-based association signals replicated.

Monocyte Count

We found mostly non-coding variants associated with monocyte counts at the exome-wide level. One exception was the rs1292053 (p.Met76Thr) missense variant in *TUBD1*, for the multi-ancestry meta-analysis (p value = 6.53×10^{-12}). Although the association was consistent across neutrophil and lymphocyte counts, the association with total WBC was almost entirely driven by the strong association with monocyte counts. An intergenic variant (rs4917014) near *C7orf72-IKZF1* (MIM: 603023) was associated with lower monocyte count in the multi-ancestry meta-analysis (p value = 9.75×10^{-12}). It was not associated with any other WBC sub-type. An intronic variant (rs11625112) in *SLC7A8* (MIM: 600749) was associated with lower monocyte counts in the EA meta-analysis (p value = 1.03×10^{-9}). We also found a secondary signal, rs11642873 near *IRF8* (MIM: 601565),[37] that was associated with higher monocyte count in the EA meta-analysis (discovery beta [p value] = 0.072 [1.40×10^{-22}], conditional beta [p value] = 0.054 [1.41×10^{-11}]). Similar to their association with monocyte count, both rs11625112 in *SLC7A8* and rs11642873 near *IRF8* had consistent associations with basophil and eosinophil counts, but were not seen in AAs and HAs.

Lymphocyte Count

An intronic variant (rs4763879) in *CD69* (MIM: 107273) was associated with decreased lymphocyte count in the EA meta-analysis (p value = 1.59×10^{-10}). None of the other sub-types showed an association with rs4763879. The signal was not observed in AAs or HAs. A secondary missense variant (rs2229094, p.Cys13Arg [c.37T>C]) in *LTA* (MIM: 153440) was associated with higher lymphocyte count in the EA meta-analysis (p value = 3.14×10^{-12}). The association was consistent across EAs and AAs, as well as for neutrophil counts, basophil counts, and for total WBC. *LTA*-rs2229094 is located near a previously reported WBC-associated SNP rs2524079 in *LOC101929772*,[6] though the LD between these variants is quite low (r^2 = 0.04). Finally, although we observed a rare, missense variant in *TRIM6* (MIM: 607564) (rs199694284, p.Val258Ala [c.773T>C], EAF in EAs = 5.25×10^{-5}, discovery p value = 7.56×10^{-8}) associated with lymphocyte counts in EAs (Table S7), the association did not replicate.

Basophil Count

In the EA meta-analysis, we identified a 3′ UTR variant (rs2295764) in *ASXL1* (MIM: 612990) associated with lower basophil count (p value = 1.46×10^{-10}). This variant was also associated with lower eosinophil and monocyte counts. The allele frequencies differed across ethnic groups and the association was not observed in AAs or HAs.

Shared Associations of WBC Loci with Disease Phenotypes

To assess the shared association between these WBC loci and immune-mediated diseases and other relevant clinical phenotypes, we performed a PheWAS in 29,722 individuals and queried published GWAS databases of autoimmune diseases including IBD, MS, RA, SLE, and T1D. The majority of WBC variants discovered by the present study were associated with multiple autoimmune diseases. PheWAS identified *TNXB* (rs185819, p.His1161Arg) associated with risk of MS and SLE (Figure 2, Table 3). In lookups of GWAS databases, after correcting for multiple testing of 16 variants and 15 inflammatory diseases (p value < 2.08×10^{-4}), disease-variant associations were additionally detected for MS (*CD69*, *TUBD1*), IBD (*GCKR*, *LTA*, *TNXB*, *IKZF1*, *TUBD1*, *ETS2-PSMG1*), SLE (*LTA*, *IRF8*, *TNXB*, *IKZF1*), RA (*TNXB*), PBC (*LTA*), and T1D (*CD69*, *TNXB*). Additional associations between immunologically relevant clinical phenotypes and WBC trait variants included selective immunoglobulin A deficiency (MIM: 137100) with *CD69* and *IKZF1* (p value < 1.90×10^{-11}) and between *IRF8* and systemic sclerosis (MIM: 181750) (p value = 2.30×10^{-12}). The inflammatory marker C-reactive protein (CRP) was strongly associated with *GCKR* and *ETS2-PSMG1* (p value < 4.00×10^{-8}) (Tables 3 and S11).

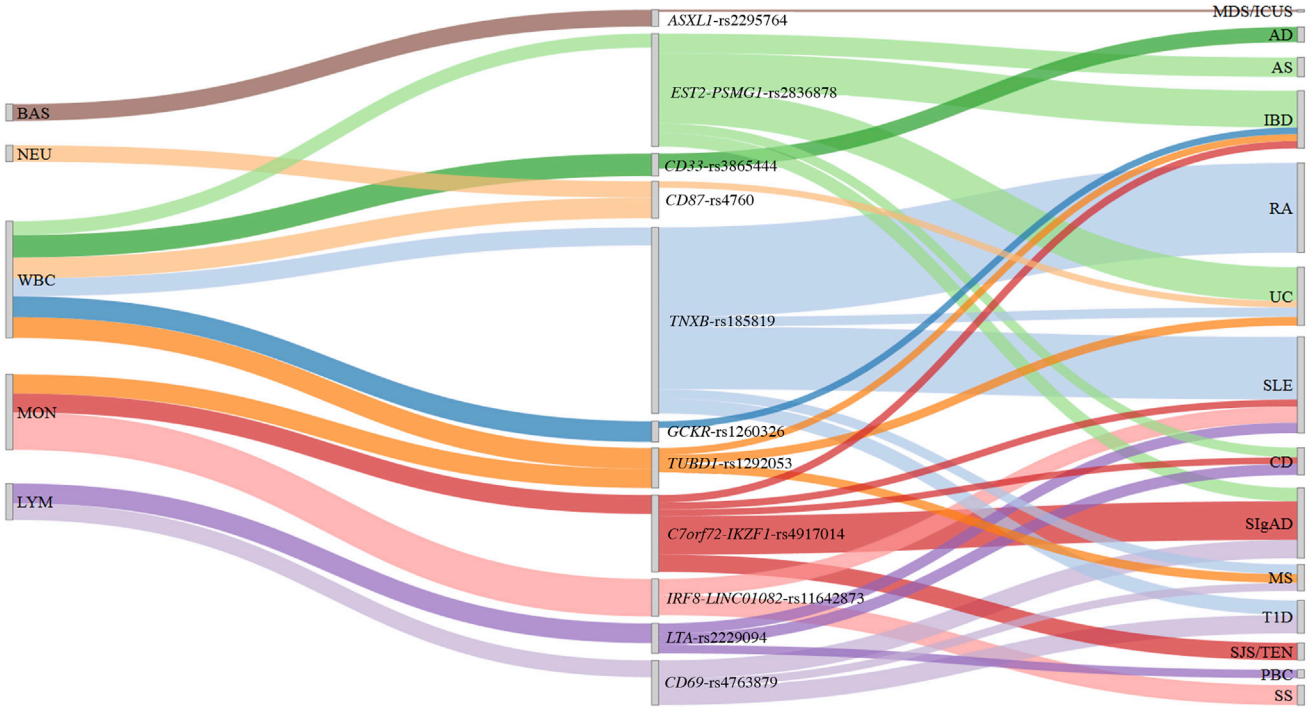


Figure 2. Pleiotropy Plot Showing Shared Genetic Loci between WBC Traits and Autoimmune Inflammatory and Other Immune-Mediated Diseases

The thickness of each line connecting genes with WBC subtypes and immune-mediated diseases corresponds to the observed strength of association in p values. p values for gene-disease associations were derived from published genome-wide association studies (see Material and Methods section for references). Abbreviations are as follows: AD, Alzheimer disease; AS, ankylosing spondylitis; BAS, basophils; CD, Crohn disease; ICUS, idiopathic cytopenia of undetermined significance; IBD, inflammatory bowel diseases; LYM, lymphocytes; MON, monocytes; MDS, myelodysplastic syndrome; MS, multiple sclerosis; NEU, neutrophils; PBC, primary biliary cirrhosis; RA, rheumatoid arthritis; SIgAD, selective immunoglobulin A deficiency; SJS/TEN, Stevens-Johnson syndrome/toxic epidermal necrolysis; SLE, systemic lupus erythematosus; SS, systemic sclerosis; T1D, type 1 diabetes mellitus; UC, ulcerative colitis, WBC, white blood cells.

Discussion

In this large-scale exome-wide association meta-analysis of WBC related traits in ~157,622 discovery and replication samples from five ancestries, we discovered 14 primary and 2 secondary SNV associations with total WBC and differential counts in EAs and the combined multi-ancestry samples, substantially increasing the number of loci associated with these hematologic traits. We observed shared genetic mechanisms influencing variations in WBC counts and susceptibility to chronic inflammatory and autoimmune diseases. These include genes and pathways involved in hematopoietic stem cell differentiation, apoptosis, cell adhesion, centrosome, and microtubule function.

Our statistical thresholds to declare significance at the discovery stage ($p < 2 \times 10^{-7}$ in the single-variant analyses) was adjusted for the approximate number of variants genotyped on the ExomeChip. Although we did not explicitly correct for testing multiple traits, the p values of our reported variants (Table 2) all pass the 5.0×10^{-8} standard of evidence for genome-wide association studies of correlated traits.[38] Furthermore, we relied on independent replication to confirm our observed associations. Despite the limited size of our replication set, it is noteworthy that we robustly replicated both known and novel WBC variants, suggesting a very low probability of reporting false-positive associations.

To quantitatively assess the contribution of loci identified by our Exomechip analysis, we have performed a comparative analysis of the proportion of total WBC phenotypic variance explained in a random sub-sample of 17,306 EAs from our largest discovery cohort, the WHI study. The proportion of variance in total WBC explained by the 28 previously known GWAS loci is 0.0137. The proportion of variance explained by the combination of known GWAS loci plus the ten additional Exomechip-identified loci we report is 0.0183. Thus, our Exomechip analysis has resulted in a 34% increase in the proportion of variance explained for total WBC in whites. These results suggest the possibility that exonic variants and/or variants not well-captured by traditional GWAS arrays may make an important contribution to the genetic architecture of WBC traits.

Loci Involving Hematopoietic Lineage Differentiation and Activation of Cell Surface Receptors

Consistent with the pattern of association of the *CD33* index SNP rs3865444 with lower total WBC count involving

Table 3. Association of White Blood Cell Trait Variants with Immune-Mediated Diseases and Clinical Phenotypes in Previous Genome-wide Association Studies

Trait (Population)	dbSNP ID	Chr	Pos	Alt/Ref	Gene	Phenotype	Sample Size	p Value[a]
WBC (EA)	rs1260328	2	27,730,940	C/T	*GCKR*	inflammatory bowel disease	96,486	1.27×10^{-4}
LYM (EA)	rs2229094[b]	6	31,540,556	C/T	*LTA*	Crohn disease	69,268	7.81×10^{-7}
LYM (EA)	rs2229094[b]	6	31,540,556	C/T	*LTA*	systemic lupus erythematosus	23,209	3.09×10^{-7}
LYM (EA)	rs2229094[b]	6	31,540,556	C/T	*LTA*	primary biliary cirrhosis	21,216	1.31×10^{-5}
WBC (All)	rs185819	6	32,050,067	C/T	*TNXB*	multiple sclerosis[c]	22,850	2.16×10^{-6}
WBC (All)	rs185819	6	32,050,067	C/T	*TNXB*	type 1 diabetes	33,394	3.29×10^{-9}
WBC (All)	rs185819	6	32,050,067	C/T	*TNXB*	ulcerative colitis	72,647	2.91×10^{-6}
WBC (All)	rs185819	6	32,050,067	C/T	*TNXB*	systemic lupus erythematosus[c]	23,209	2.32×10^{-37}
WBC (All)	rs185819	6	32,050,067	C/T	*TNXB*	rheumatoid arthritis	103,558	3.90×10^{-53}
MON (All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	systemic lupus erythematosus	32,444	8.10×10^{-5}
MON (All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	inflammatory bowel disease	96,486	4.59×10^{-5}
MON (All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	Crohn disease	69,268	1.49×10^{-4}
MON (All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	Stevens-Johnson syndrome/ toxic epidermal necrolysis	1,129	8.00×10^{-11}
MON (All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	selective immunoglobulin a deficiency	2,748	2.80×10^{-23}
LYM (EA)	rs4763879	12	9,910,164	A/G	*CD69*	type 1 diabetes	38,522	1.90×10^{-11}
LYM (EA)	rs4763879	12	9,910,164	A/G	*CD69*	multiple sclerosis	38,135	2.18×10^{-5}
LYM (EA)	rs4763879	12	9,910,164	A/G	*CD69*	selective immunoglobulin A deficiency	2,748	1.90×10^{-11}
MON (EA)	rs11642873	16	85,991,705	C/A	*IRF8-LINC01082*	systemic lupus erythematosus	23,209	3.56×10^{-10}
MON (EA)	rs11642873	16	85,991,705	C/A	*IRF8-LINC01082*	systemic sclerosis	14,853	2.30×10^{-12}
WBC (EA); MON (All)	rs1292053	17	57,963,537	G/A	*TUBD1*	multiple sclerosis	38,135	7.47×10^{-6}
WBC (EA); MON (All)	rs1292053	17	57,963,537	G/A	*TUBD1*	Crohn disease	96,486	8.53×10^{-6}
WBC (EA); MON (All)	rs1292053	17	57,963,537	G/A	*TUBD1*	inflammatory bowel disease	96,486	9.61×10^{-5}
WBC (EA); NEU (EA)	rs4760	19	44,153,100	G/A	*CD87* (*PLAUR*)	ulcerative colitis	72,647	1.51×10^{-4}
WBC (EA)	rs3865444	19	51,727,962	A/C	*CD33*	Alzheimer disease	59,716	1.60×10^{-9}
BAS (EA)	rs2295764	20	31,025,163	G/A	*ASXL1*	somatic mutations in MDS, CML, and ICUS	–	–
WBC (All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	ankylosing spondylitis	9,609	4.90×10^{-12}
WBC (All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	Crohn disease	69,268	2.43×10^{-6}
WBC (All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	ulcerative colitis	72,647	2.05×10^{-20}
WBC (All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	inflammatory bowel disease	96,486	3.70×10^{-22}
WBC (All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	selective immunoglobulin A deficiency	2,748	1.40×10^{-8}

Abbreviations: Chr, chromosome; Pos, basepair position; Alt, effect allele; Ref, reference allele; CML, chronic myelogenous leukemia; ICUS, Idiopathic cytopenia of undetermined significance; MDS, myelodysplastic syndrome, WBC, white blood cell; NEU, neutrophil; MON, monocyte; LYM, lymphocyte; BAS, basophil.
[a]Significant results are shown after correcting for multiple testing of 16 variants and 15 diseases ($p < 2.08 \times 10^{-4}$). When multiple studies report the same variant-trait associations, results from the largest sample size are presented here.
[b]LD r^2 between rs2229094 and rs1799964 is 0.75.
[c]Phenome-wide association results. Permutation p value for association with multiple sclerosis was 0.0122.

all myeloid lineages (and lower platelet count) (Table S12), CD33 is an early myeloid differentiation antigen and cell surface receptor that binds sialic acid-containing ligands and mediates diverse inhibitory functions of WBC in the innate immune system.[39] *CD33* is also highly expressed on the surface of acute myeloid leukemia (AML) cells. *CD33* rs3865444 is in complete LD with *CD33* rs12459419 (p.Ala14Val), the presumed functional variant that results in lower full-length *CD33* expression due to skipping of exon 2.[40]

PLAUR encodes for the glycosyl-phosphatidylinositol-anchored urokinase plasminogen activator receptor (*UPAR*). UPAR, also known as CD87, is a differentiation antigen on cells of the myelomonocytic lineage and also an activation antigen on monocytes and T lymphocytes.[41,42] The deleterious coding variant of *CD87* rs4760A>G (p.Leu272Pro)[36] is also a strong eQTL for *CD87* expression in monocytes and whole blood (Table S13). In addition to its role in plasminogen activation and fibrinolysis, UPAR is involved in cell adhesion, migration, and chemotaxis and is a regulator of the uptake by macrophages of apoptotic neutrophils.[43] It is possible that the latter mechanism might explain the selective association of rs4760 with lower neutrophil count.

The *CD69* intronic allele rs4763879G>A was associated with lower lymphocyte count but not with other WBC types. Accordingly, *CD69* encodes a calcium-dependent lectin superfamily of type II transmembrane cell surface receptor involved in regulation of lymphocyte proliferation.[44] As an early activation marker of lymphocytes, CD69 inhibits egress of lymphocytes into the circulation by downregulating sphingosine-1-phosphate receptor type 1 (MIM: 601974).[44] Notably, *CD69* rs4763879 correlates with the expression of *CD69* in monocytes and with the expression of C-type lectin domain family member genes *CLECL1* and *CLEC2D* in lymphoid cells (Table S13).

The intronic variant rs9374080 of non-coding RNA/pseudogene *CCDC162P* has been previously associated with red blood cell traits—lower mean corpuscular volume, mean corpuscular hemoglobin[45]—and with platelet traits (Tables S11 and S12). In this study, we extend the association of rs9374080 to higher total WBC and myeloid-derived cell counts, including basophil count (Table S8). The index SNP is located ~70 kb 3′ of *CD164* (endolyn) (MIM: 603356), which encodes a small transmembrane sialomucin protein on the surface of early hematopoietic progenitors, maturing erythroid cells, and activated basophils.[46] *CD164* regulates CXCR4/CXCL12 signaling in hematopoietic precursor cells.[47] The region of association is located within a putative regulatory region enriched in epigenomic marks and ChIP-seq sites for various hematopoietic transcription factors (*GATA1*, *TAL1*) in K562 erythroleukemia and lymphoblastoid cell lines.[48] These observations fit with the broad pattern of association of this variant with multiple blood cell lineages.

Loci Involving Hematopoietic Transcription Factors and Epigenetic Modifiers

We identified variants in or near multiple genes encoding hematopoietic transcription factors that are associated with WBC traits. These loci include *IRF8-LINC01082*, *C7orf72-IKZF1*, *SLC7A8-CEBPE*, *JMJD1C*, *ASXL1*, and *ETS2-PSMG1*.

The 3′ UTR variant rs2295764 of *ASXL1*, which was significantly associated with *ASXL1* transcript expression, was associated with lower basophil count and to a lesser degree with lower monocyte and eosinophil counts and also to some extent with higher red cell distribution width (Tables S12 and S13). *ASXL1* is a chromatin binding transcriptional regulator of the polycomb group and hematopoietic tumor suppressor gene.[49] *JMJD1C* is also an epigenetic regulator of gene expression, probably through histone demethylation.[50] The association between *JMJD1C* and lower WBC counts (this study), platelet count, mean platelet volume, and platelet reactivity[51] indicate multi-lineage effects on hematopoiesis. *JMJD1C* was originally identified as a ligand-dependent interacting partner of thyroid hormone and androgen receptors.[52] In human myeloid leukemia cells, *JMJD1C* functions as a coactivator for the leukemogenic transcriptional complex *RUNX1-RUNX1T1* to increase AML cell proliferation and survival.[53] An intergenic variant rs2836878 located between *ETS2* and *PSMG1* showed evidence of multi-lineage association with lower total WBC count across all myeloid cell types and to a lesser extent with lower platelet count and higher hemoglobin (Table S12); rs2836878 is a whole-blood eQTL for *ETS2* (Table S13). *ETS2* is another proto-oncogene that encodes for a transcription factor involved in stem cell development, cell senescence, and death, whereas the product of *PSMG1* is involved in maturation of proteasomes. *ETS2*, which is highly expressed in monocytes but not in granulocytes, has been shown to be involved in macrophage differentiation, regulation of megakaryocytic gene expression, T cell development, and phenotypic switch from erythroid to megakaryocytic development in hematopoietic cells.[54]

We identified several variants associated with monocyte count in loci that involve hematopoietic transcription factor genes (*IRF8*, *SLC7A8*, and *IKZF1*), further supporting their role in regulation of myelopoiesis and granulocyte/monocyte lineage fate. The minor C allele of rs11642873, located 35 kb 3′ of *IRF8*, was associated with higher monocyte count (and to a lesser degree with higher eosinophil and basophil counts) (Table S8). In eQTL analysis, an *IRF8* variant rs17445836 is in moderate LD with the *IRF8* rs11642873 variant ($r^2 = 0.48$) that has a *cis*-regulatory effect on *IRF8* expression in $CD14^+$ monocytes (Table S13). *IRF8* encodes a transcription factor critical for myeloid lineage commitment by promoting differentiation of monocytes/dendritic cells and suppressing granulopoiesis.[55] *Irf8*$^{-/-}$ mice have a myeloproliferative disorder with markedly increased number of macrophages and granulocytes in bone marrow, spleen, and lymph nodes as well as increased number of granulocytes in peripheral blood, suggesting a tumor-suppressive role of *IRF8*.[56]

Another non-coding variant associated with lower monocyte count, and to a lesser extent with lower basophil and eosinophil counts, was the intronic variant rs11625112 of *SLC7A8*, which encodes an amino acid transporter highly expressed in absorptive epithelia of the kidney and small intestine and also in the brain.[57] The index SNP is located within a blood cell DNase hypersensitivity site ~8 kb upstream of *CEBPE*, which encodes a

hematopoietic transcription factor essential for terminal differentiation and functional maturation of granulocytes.[58] Recent data also suggest a role of *CEBPE* isoforms in differential regulation of eosinophil production as well as in the monocyte-granulocyte lineage decision.[59]

The transcription factor encoded by *IKZF1* or Ikaros was initially described as a regulator of lymphoid lineage differentiation and hematopoietic progenitor cell self-renewal.[60] An Ikaros isoform selectively expressed in myeloid precursor cells was subsequently found to regulate myeloid differentiation.[60] The minor allele of intergenic variant rs4917014 in *C7orf72-IKZF1* associated selectively with lower monocyte count is located ~50 kb upstream of *IKZF1* within an LD block enriched in hematopoietic cell DNase hypersensitivity sites and enhancer histone markers, several of which are also located within ChIP-seq binding sites for the myeloid transcription factor *PU.1*.[48] The index SNP is also a monocyte and whole blood *trans*-eQTL for several immune response genes (Table S13). Further studies are required to assess whether the upstream *IKZF1* or *CEBPE* regulatory elements harboring the index SNP are important for isoform- or lineage-specific monocyte development.

Loci Involved in Regulation of Cell Death and Apoptosis

Apoptosis regulates hematopoietic stem cells and maintains the balance between cell proliferation and cell death.[61] Altered apoptotic processes contribute to the development of autoimmune and other inflammatory diseases.[62] We identified associations between WBC traits and coding variants in two additional genes involved in apoptosis. *C10orf54* rs3747869 (p.Asp187Glu) was associated with higher total WBC and neutrophil counts. The product of *C10orf54* (also known as Death Domain 1-alpha, *DD1a*), a direct transcriptional target of p53, regulates apoptosis and clearance of apoptotic cells, processes that are critical for resolution of inflammation, immune tolerance, and regulation of autoimmune responses.[63] *DD1a* is exclusively expressed within the hematopoietic compartment (monocytes, mature T cells, and macrophages) and functions as a negative immune checkpoint regulator for T cell activation and response.[64]

LTA rs2229094 (p.Cys13Arg) was associated with higher lymphocyte count. *LTA* encodes a member of the tumor necrosis factor family involved in lymphoid organ development and apoptosis.[65] Loss of *LTA* was associated with a 4-fold increase in B lymphocytes in peripheral blood count in mice.[65] The index missense SNP is also a *cis*-eQTL for *LTA* and *NFKBIL1* (MIM: 601022) (Table S13).

Loci Involved in Other Cellular and Inflammatory Processes

We identified several missense variants (*TNXB* rs185819 [p.His1161Arg], *TUBD1* rs1292053 [p.Met76Thr], and *KIF9* rs2276853 [p.Arg573Trp]) in genes involved in other cellular processes that might be relevant to WBC production or immune function. *TNXB* encodes a member of the tenascin family of extracellular matrix glycoproteins and inhibits cell adhesion and migration.[66] The index SNP localizes to the major histocompatibility complex class III region on chromosome 6 and overlaps *ATF6B* and *CYP21A2* at its 5′ and 3′ ends, respectively. The missense SNP is also an eQTL in blood or lymphoblastoid cell lines for several class II HLA genes (Table S13). The pattern of association of *TNXB* rs185819 suggests an effect at an early stage of myeloid and lymphoid differentiation. *ATF6B*, a member of the ATF6-related family of transcription factors that operate in the unfolded protein response,[67] is also a key virulence factor for *Toxoplasma gondii*.[68]

Although the role of *TUBD1* and *KIF9* on hematopoiesis is not known, both genes are involved in the structure and function of microtubules and centrosomes that are important for cell division and proliferation.[69] *TUBD1* encodes for delta-tubulin microtubule protein that is associated with centrosome structure and function. The *TUBD1* rs1292053 (p.Met76Thr), which was associated with both total WBC and monocyte counts, and to some extent with red cell and platelet parameters, is in LD with a number of SNPs in neighboring genes *RPS6KB1* and *RNFT1* and is a blood eQTL for *RNFT1* (Tables S12 and S13). *RPS6KB1* encodes a member of the ribosomal S6 kinase family of serine/threonine kinases and is part of the PI3K/AKT/mTOR signaling pathway that plays a central role in a wide spectrum of cellular activities, including cell proliferation, survival, and differentiation.[70] The PI3K pathway is also involved in Toll-like receptor (TLR) signaling and release of cytokines from macrophages,[71] and a proxy SNP of *TUBD1* rs1292053 has been associated with CRP.[72] *KIF9* is a member of the kinesin family of genes related to microtubule binding and microtubule motor activity. The *KIF9* rs2276853 variant is in LD with about 50 other variants spanning two other genes, *SETD2* and *KLHL18*, several of which are within epigenomic blood cell marks and eQTLs for *KIF9*, *KLHL18*, and *NBEAL2*.[48]

The *GCKR* rs1260326 variant is an eQTL for *SNX17*, which has been associated with T cell activation and is a binding protein for human papillomavirus L2 capsid protein and for *NRBP1*, which binds a Dengue virus protein.[48,73]

Relationship of WBC Loci to Autoimmune and Chronic Inflammatory Diseases

Abnormal immune response by lymphocytes and other white blood cells directed against self-antigens can lead to tissue injury and development of autoimmune diseases.[74] Our results add to recent evidence that genetic factors controlling WBC and immune cell counts contribute to autoimmune disease risk.[75] Several loci involve regulation of cellular mechanisms critical in the development of autoimmune diseases such as modulation of autoimmune reactivity (*CD69*)[76] and apoptosis (*LTA*, *DD1a*, *CD87*).[43,63]

The majority of our WBC-associated loci that showed substantial overlap were also associated with risk of various autoimmune and inflammatory diseases including IBD, RA, SLE, T1D, PBC, systemic sclerosis, Alzheimer disease, and Stevens-Johnson syndrome (Figure 2, Table 3). Although many of these genetic susceptibility loci are shared between different autoimmune diseases, other loci appear to be more restricted to particular cellular contexts. For example, there is an over-representation of SLE loci expressed selectively in B cells; RA-associated loci are preferentially expressed in $CD4^+$ effector T memory cells; epithelial-associated stimulated dendritic cell genes in Crohn disease; and monocyte-specific eQTLs among neurodegenerative disease variants.[77,78]

Abnormal inflammatory response and activation of microglial cells are linked with the development of AD and other neurodegenerative diseases. The WBC-associated gene *CD33* is among the inflammation-related AD risk loci identified by GWASs.[79] A variant in this gene was shown to modulate *CD33* exon 2 splicing efficiency, leading to abnormal activation of microglial cells that are tissue-resident macrophages of the brain derived from monocyte lineage cells.[79] In eQTL analysis of neuropathologically normal human brain tissues, *CD33* rs3865444 is a *cis*-eQTL of C-type lectin domain family 11 member A (*CLEC11A*) that functions as growth factor for hematopoietic progenitor cells.[80] Several of the same loci are involved in susceptibility to infectious diseases (*IRF8* and mendelian susceptibility to mycobacterial disease [MIM: 209950],[81] *TNXB* associated with *T. gonadii* and climatic adaptation,[68,82] malaria with *ABO* [MIM: 110300] and *DARC*,[3,83] *CD87* with clearance of bacteria[84]), highlighting the evolutionary trade-offs between protection against pathogens and risk of chronic disease later in life.

Relationship of WBC Loci to Hematologic Disease and Therapy

Hematopoiesis is controlled by the differential expression of key transcription factors that act cooperatively to maintain a well-orchestrated balance of hematopoietic stem cell self-renewal and differentiation.[85] These functions of transcription factors are frequently dysregulated in leukemia by chromosomal translocations, mutations, or aberrant expression and lead to abnormal self-renewal. Several of the WBC loci have additional relationships to hematologic disease and therapeutics. *CD33* is expressed in the brain and on AML blasts and leukemic stem cells and has therefore been exploited therapeutically as a target for anti-leukemic therapy.[40] The *CD33* rs3865444 and rs12459419 variants associated with lower WBC count and alternative splicing of exon 2, respectively, have been associated with both Alzheimer disease risk and AML treatment efficacy.[40] The exon 2 region of *CD33* is important for sialic acid binding, microglial cell phagocytosis of beta-amyloid, and an epitope recognized by the antibody-targeted chemotherapy agent gemtuzumab ozogamicin.[40,86] *CD87* is expressed on various immune cells including neutrophils, monocytes, macrophages, T cells, and basophils, as well as endothelial cells and hepatocytes.[41,87] The cleaved soluble form of CD87 might have a role in hematopoietic stem/progenitor cell mobilization.[88]

Somatic mutations in *ASXL1* are associated with risk of myelodysplastic syndrome (MDS [MIM: 614286]), chronic myelomonocytic leukemia (CMML [MIM: 607785]), and idiopathic cytopenia of undetermined significance (ICUS).[49,89,90] Knockdown of *Asxl1* in mouse results in impaired lymphoid and myeloid differentiation and multi-lineage cytopenias.[91] Collectively, these results suggest that both germline and somatic mutations in *ASXL1* cause lower blood cell counts. The transcription factor *ETS2* has been shown to regulate phenotypic switch from erythroid to megakaryocyte in acute megakaryocytic leukemia (AMKL), and overexpression of *ETS2* results in altered sensitivity to chemotherapy drugs.[54] Recent studies have shown that *IKZF1* deletions and mutations that caused reduction of Ikaros activity are highly associated with development of acute lymphoblastic leukemia.[92,93] On the other hand, depletion of *JMJD1C* leads to growth impairment of a variety of leukemic cell types without noticeable effects on normal hematopoietic cells.[52] Therefore, *JMJD1C* is a potentially relevant drug target for leukemia.

Besides the single-variant association results, we confirmed previously reported gene-based association results for WBC count (*CXCR2*)[12] and monocytes (*IL17RA*) (N.P., U.M.S., J.B.-J., and M.-H.C., unpublished data). We also identified an additional gene putatively associated with WBC count (*TAF3*). *IL17RA* is widely expressed in myelomonocytic cells, lymphocytes, and bone marrow stromal cells and is part of the IL-17 cytokine signaling pathway that plays role in hematopoiesis, promotes inflammation, and is implicated in autoimmune diseases such as psoriasis, RA, and IBD.[94] *TAF3*, which encodes for a TATA-box binding protein, is located near *GATA3*, a transcription factor important for T lymphocyte differentiation. Variants in *TAF3* are associated with mean corpuscular hemoglobin concentration[95] whereas *GATA3* variants have been associated with susceptibility to hematologic malignancies.[96] Despite our large sample size, power to detect rare variants of more modest effect, either individually or aggregated into gene-based tests, may be limited. Future studies will require enormous sample sizes, probably considerably larger than in the current study, in order to detect additional rare variants (both individually and in aggregate) of moderate effect sizes associated with complex traits.

Our study has both strengths and limitations. By combining data from 25 studies world-wide, we were able to investigate the effect sizes and allele frequencies of variants in multiple ancestry groups. Variants with consistent effects across ancestries serve as strong candidates for causal variants. In addition to our ability to investigate how genetic variants influence WBC sub-types, our discovery analyses were well powered to detect moderate effect

sizes. Indeed, although we did not correct for testing seven different phenotypes in three different meta-analyses, the combined p values of our reported variants (Table 2) all pass the 5.0×10^{-8} standard of evidence for genome-wide association studies of correlated traits.[38] We note that some cohorts did not measure a differential WBC in addition to total WBC, which limited our ability to assess associations with specific WBC subtypes in some instances.

In conclusion, by combining WBC exome-array analysis with PheWAS and functional annotation of variants, we identified likely causal variants associated with total and differential WBC counts as well as risk of autoimmune and inflammatory diseases. These results shed light on genetic mechanisms that regulate WBC counts and suggest a shared genetic architecture with predisposition to autoimmune and chronic inflammatory diseases. Future studies in model organisms are required to elucidate the underlying molecular mechanisms of how these genes result in variations in WBC count and development of autoimmune diseases.

Supplemental Data

Supplemental Data include 2 figures, 13 tables, and additional funding information and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.05.003.

Acknowledgments

We thank all participants and study coordinating centers of the participating studies and cohorts. P.L.A. was supported by NHLBI R21 HL121422-02. G.L. was supported by the Canada Research Chair program and the Canadian Institute of Health Research MOP#123382. This work was supported in part by the National Institute on Aging, NIH Intramural Research Program. Data analyses utilized the computational resources of the NIH HPC Biowulf cluster at the NIH. The Framingham Heart Study (FHS) acknowledges the Shared Computing Cluster at Boston University. Infrastructure and data analysis for the ARIC study was partly supported by Grant Number UL1RR025005, a component of the NIH Roadmap for Medical Research, and grant R01 HL086694 from the NHLBI. Airwave study thanks Louisa Cavaliero who assisted in data collection and management, Peter McFarlane and the Glasgow CARE, Patricia Munroe at Queen Mary University of London, and Joanna Sarnecka and Ania Zawodniak at Northwick Park for their contributions to the study. SOLID-TIMI-52 and STABILITY thank Liling Warren for contributions to the genetic analysis of the study datasets. Estonian Genome Center, University of Tartu (EGCUT) thanks Mr. V. Soo, Mr. S. Smith, and Dr. L. Milani for their contribution. The Rotterdam Study (RS) thanks Ms. Mila Jhamai, Ms. Sarah Higgins, and Mr. Marijn Verkerk for their help in creating the Exomechip database, and Ms. Carolina Medina-Gomez, Mr. Lennard Karsten, and Dr. Linda Broer. Additional acknowledgments and funding information are provided in the Supplemental Data.

Received: February 17, 2016
Accepted: May 3, 2016
Published: June 23, 2016

Web Resources

1000 Genomes, http://www.1000genomes.org
BCX ExomeChip association results, http://www.mhi-humangenetics.org/en/resources
CheckVCF, https://github.com/zhanxw/checkVCF
GRASP, http://grasp.nhlbi.nih.gov/Overview.aspx
HPC @ NIH, http://hpc.nih.gov
OMIM, http://www.omim.org/
R statistical software, http://www.r-project.org/
RareMETALS, http://genome.sph.umich.edu/wiki/RareMETALS
RareMetalWorker, http://genome.sph.umich.edu/wiki/RAREMETALWORKER
RvTests, http://genome.sph.umich.edu/wiki/RvTests

References

1. Chen, L., Kostadima, M., Martens, J.H., Canu, G., Garcia, S.P., Turro, E., Downes, K., Macaulay, I.C., Bielczyk-Maczynska, E., Coe, S., et al.; BRIDGE Consortium (2014). Transcriptional diversity during lineage commitment of human blood progenitors. Science *345*, 1251033.
2. Gasteiger, G., and Rudensky, A.Y. (2014). Interactions between innate and adaptive lymphocytes. Nat. Rev. Immunol. *14*, 631–639.
3. Reiner, A.P., Lettre, G., Nalls, M.A., Ganesh, S.K., Mathias, R., Austin, M.A., Dean, E., Arepalli, S., Britton, A., Chen, Z., et al. (2011). Genome-wide association study of white blood cell count in 16,388 African Americans: the continental origins and genetic epidemiology network (COGENT). PLoS Genet. *7*, e1002108.
4. Kong, M., and Lee, C. (2013). Genetic associations with C-reactive protein level and white blood cell count in the KARE study. Int. J. Immunogenet. *40*, 120–125.
5. Crosslin, D.R., McDavid, A., Weston, N., Nelson, S.C., Zheng, X., Hart, E., de Andrade, M., Kullo, I.J., McCarty, C.A., Doheny, K.F., et al.; Electronic Medical Records and Genomics (eMERGE) Network (2012). Genetic variants associated with the white blood cell count in 13,923 subjects in the eMERGE Network. Hum. Genet. *131*, 639–652.
6. Nalls, M.A., Couper, D.J., Tanaka, T., van Rooij, F.J., Chen, M.H., Smith, A.V., Toniolo, D., Zakai, N.A., Yang, Q., Greinacher, A., et al. (2011). Multiple loci are associated with white blood cell phenotypes. PLoS Genet. *7*, e1002113.
7. Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y., and Kamatani, N. (2010). Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat. Genet. *42*, 210–215.
8. Keller, M.F., Reiner, A.P., Okada, Y., van Rooij, F.J., Johnson, A.D., Chen, M.H., Smith, A.V., Morris, A.P., Tanaka, T., Ferrucci, L., et al.; CHARGE Hematology; COGENT; BioBank Japan Project (RIKEN) Working Groups (2014). Trans-ethnic meta-analysis of white blood cell phenotypes. Hum. Mol. Genet. *23*, 6944–6960.
9. Soranzo, N., Spector, T.D., Mangino, M., Kühnel, B., Rendon, A., Teumer, A., Willenborg, C., Wright, B., Chen, L., Li, M., et al. (2009). A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat. Genet. *41*, 1182–1190.
10. Li, J., Glessner, J.T., Zhang, H., Hou, C., Wei, Z., Bradfield, J.P., Mentch, F.D., Guo, Y., Kim, C., Xia, Q., et al. (2013). GWAS of blood cell traits identifies novel associated loci and epistatic

interactions in Caucasian and African-American children. Hum. Mol. Genet. *22*, 1457–1464.

11. Auer, P.L., Johnsen, J.M., Johnson, A.D., Logsdon, B.A., Lange, L.A., Nalls, M.A., Zhang, G., Franceschini, N., Fox, K., Lange, E.M., et al. (2012). Imputation of exome sequence variants into population- based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO Exome Sequencing Project. Am. J. Hum. Genet. *91*, 794–808.
12. Auer, P.L., Teumer, A., Schick, U., O'Shaughnessy, A., Lo, K.S., Chami, N., Carlson, C., de Denus, S., Dubé, M.P., Haessler, J., et al. (2014). Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. Nat. Genet. *46*, 629–634.
13. Nalls, M.A., Wilson, J.G., Patterson, N.J., Tandon, A., Zmuda, J.M., Huntsman, S., Garcia, M., Hu, D., Li, R., Beamer, B.A., et al. (2008). Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. Am. J. Hum. Genet. *82*, 81–87.
14. Grove, M.L., Yu, B., Cochran, B.J., Haritunians, T., Bis, J.C., Taylor, K.D., Hansen, M., Borecki, I.B., Cupples, L.A., Fornage, M., et al. (2013). Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. PLoS ONE *8*, e68095.
15. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Mägi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E., et al.; Genetic Investigation of Anthropometric Traits (GIANT) Consortium (2014). Quality control and conduct of genome-wide association meta-analyses. Nat. Protoc. *9*, 1192–1212.
16. Auer, P.L., Reiner, A.P., and Leal, S.M. (2016). The effect of phenotypic outliers and non-normality on rare-variant association testing. Eur. J. Hum. Genet. Published online January 6, 2016. http://dx.doi.org/10.1038/ejhg.2015.270.
17. Liu, D.J., Peloso, G.M., Zhan, X., Holmen, O.L., Zawistowski, M., Feng, S., Nikpay, M., Auer, P.L., Goel, A., Zhang, H., et al. (2014). Meta-analysis of gene-level tests for rare variant association. Nat. Genet. *46*, 200–204.
18. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
19. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
20. Pulley, J., Clayton, E., Bernard, G.R., Roden, D.M., and Masys, D.R. (2010). Principles of human subjects protections applied in an opt-out, de-identified biobank. Clin. Transl. Sci. *3*, 42–48.
21. Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat. Biotechnol. *31*, 1102–1110.
22. Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al.; CARDIoGRAMplusC4D Consortium (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. *47*, 1121–1130.
23. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al.; International Multiple Sclerosis Genetics Consortium; International IBD Genetics Consortium (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986.
24. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al.; RACI consortium; GARNET consortium (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381.
25. Bentham, J., Morris, D.L., Cunninghame Graham, D.S., Pinder, C.L., Tombleson, P., Behrens, T.W., Martín, J., Fairfax, B.P., Knight, J.C., Chen, L., et al. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat. Genet. *47*, 1457–1464.
26. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E., et al.; International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2 (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature *476*, 214–219.
27. Tsoi, L.C., Spain, S.L., Knight, J., Ellinghaus, E., Stuart, P.E., Capon, F., Ding, J., Li, Y., Tejasvi, T., Gudjonsson, J.E., et al.; Collaborative Association Study of Psoriasis (CASP); Genetic Analysis of Psoriasis Consortium; Psoriasis Association Genetics Extension; Wellcome Trust Case Control Consortium 2 (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat. Genet. *44*, 1341–1348.
28. Bradfield, J.P., Qu, H.Q., Wang, K., Zhang, H., Sleiman, P.M., Kim, C.E., Mentch, F.D., Qiu, H., Glessner, J.T., Thomas, K.A., et al. (2011). A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet. *7*, e1002293.
29. Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A., Adány, R., Aromaa, A., et al. (2010). Multiple common variants for celiac disease influencing immune gene expression. Nat. Genet. *42*, 295–302.
30. Cordell, H.J., Han, Y., Mells, G.F., Li, Y., Hirschfield, G.M., Greene, C.S., Xie, G., Juran, B.D., Zhu, D., Qian, D.C., et al.; Canadian-US PBC Consortium; Italian PBC Genetics Study Group; UK-PBC Consortium (2015). International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. Nat. Commun. *6*, 8019.
31. Leslie, R., O'Donnell, C.J., and Johnson, A.D. (2014). GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. Bioinformatics *30*, i185–i194.
32. Chami, N., Chen, M.-H., Slater, A.J., Eicher, J.D., Evangelou, E., Tajuddin, S.M., Love-Gregory, L., Kacprowski, T., Schick, U.M., Nomura, A., et al. (2016). Exome genotyping identifies pleiotropic variants associated with red blood cell traits. Am. J. Hum. Genet. *99*. Published online June 23, 2016. http://dx.doi.org/10.1016/j.ajhg.2016.05.007.
33. Eicher, J.D., Chami, N., Kacprowski, T., Nomura, A., Chen, M.-H., Yanek, L.R., Tajuddin, S.M., Schick, U.M., Slater, A.J., Pankratz, N., et al. (2016). Platelet-related variants

identified by exomechip meta-analysis in 157,293 individuals. Am. J. Hum. Genet. *99*. Published online June 23, 2016. http://dx.doi.org/10.1016/j.ajhg.2016.05.005.

34. Zhang, X., Gierman, H.J., Levy, D., Plump, A., Dobrin, R., Goring, H.H., Curran, J.E., Johnson, M.P., Blangero, J., Kim, S.K., et al. (2014). Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. BMC Genomics *15*, 532.
35. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.
36. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
37. Crosslin, D.R., McDavid, A., Weston, N., Zheng, X., Hart, E., de Andrade, M., Kullo, I.J., McCarty, C.A., Doheny, K.F., Pugh, E., et al.; CHARGE Hematology Working Group; electronic Medical Records and Genomics (eMERGE) Network (2013). Genetic variation associated with circulating monocyte count in the eMERGE Network. Hum. Mol. Genet. *22*, 2119–2127.
38. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al.; Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.
39. Ulyanova, T., Blasioli, J., Woodford-Thomas, T.A., and Thomas, M.L. (1999). The sialoadhesin CD33 is a myeloid-specific inhibitory receptor. Eur. J. Immunol. *29*, 3440–3449.
40. Malik, M., Chiles, J., 3rd, Xi, H.S., Medway, C., Simpson, J., Potluri, S., Howard, D., Liang, Y., Paumi, C.M., Mukherjee, S., et al. (2015). Genetics of CD33 in Alzheimer's disease and acute myeloid leukemia. Hum. Mol. Genet. *24*, 3557–3570.
41. Elghetany, M.T., Patel, J., Martinez, J., and Schwab, H. (2003). CD87 as a marker for terminal granulocytic maturation: assessment of its expression during granulopoiesis. Cytometry B Clin. Cytom. *51*, 9–13.
42. Nykjaer, A., Møller, B., Todd, R.F., 3rd, Christensen, T., Andreasen, P.A., Gliemann, J., and Petersen, C.M. (1994). Urokinase receptor. An activation antigen in human T lymphocytes. J. Immunol. *152*, 505–516.
43. Park, Y.J., Liu, G., Tsuruta, Y., Lorne, E., and Abraham, E. (2009). Participation of the urokinase receptor in neutrophil efferocytosis. Blood *114*, 860–870.
44. Shiow, L.R., Rosen, D.B., Brdicková, N., Xu, Y., An, J., Lanier, L.L., Cyster, J.G., and Matloubian, M. (2006). CD69 acts downstream of interferon-alpha/beta to inhibit S1P1 and lymphocyte egress from lymphoid organs. Nature *440*, 540–544.
45. van der Harst, P., Zhang, W., Mateo Leach, I., Rendon, A., Verweij, N., Sehmi, J., Paul, D.S., Elling, U., Allayee, H., Li, X., et al. (2012). Seventy-five genetic loci influencing the human red blood cell. Nature *492*, 369–375.
46. Zannettino, A.C., Bühring, H.J., Niutta, S., Watt, S.M., Benton, M.A., and Simmons, P.J. (1998). The sialomucin CD164 (MGC-24v) is an adhesive glycoprotein expressed by human hematopoietic progenitors and bone marrow stromal cells that serves as a potent negative regulator of hematopoiesis. Blood *92*, 2613–2628.
47. Forde, S., Tye, B.J., Newey, S.E., Roubelakis, M., Smythe, J., McGuckin, C.P., Pettengell, R., and Watt, S.M. (2007). Endolyn (CD164) modulates the CXCL12-mediated migration of umbilical cord blood CD133+ cells. Blood *109*, 1825–1833.
48. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. *40*, D930–D934.
49. Gelsi-Boyer, V., Trouplin, V., Adélaïde, J., Bonansea, J., Cervera, N., Carbuccia, N., Lagarde, A., Prebet, T., Nezri, M., Sainty, D., et al. (2009). Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br. J. Haematol. *145*, 788–800.
50. Watanabe, S., Watanabe, K., Akimov, V., Bartkova, J., Blagoev, B., Lukas, J., and Bartek, J. (2013). JMJD1C demethylates MDC1 to regulate the RNF8 and BRCA1-mediated chromatin response to DNA breaks. Nat. Struct. Mol. Biol. *20*, 1425–1433.
51. Johnson, A.D., Yanek, L.R., Chen, M.H., Faraday, N., Larson, M.G., Tofler, G., Lin, S.J., Kraja, A.T., Province, M.A., Yang, Q., et al. (2010). Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. Nat. Genet. *42*, 608–613.
52. Sroczynska, P., Cruickshank, V.A., Bukowski, J.P., Miyagi, S., Bagger, F.O., Walfridsson, J., Schuster, M.B., Porse, B., and Helin, K. (2014). shRNA screening identifies JMJD1C as being required for leukemia maintenance. Blood *123*, 1870–1882.
53. Chen, M., Zhu, N., Liu, X., Laurent, B., Tang, Z., Eng, R., Shi, Y., Armstrong, S.A., and Roeder, R.G. (2015). JMJD1C is required for the survival of acute myeloid leukemia by functioning as a coactivator for key transcription factors. Genes Dev. *29*, 2123–2139.
54. Ge, Y., LaFiura, K.M., Dombkowski, A.A., Chen, Q., Payton, S.G., Buck, S.A., Salagrama, S., Diakiw, A.E., Matherly, L.H., and Taub, J.W. (2008). The role of the proto-oncogene ETS2 in acute megakaryocytic leukemia biology and therapy. Leukemia *22*, 521–529.
55. Yáñez, A., Ng, M.Y., Hassanzadeh-Kiabi, N., and Goodridge, H.S. (2015). IRF8 acts in lineage-committed rather than oligopotent progenitors to control neutrophil vs monocyte production. Blood *125*, 1452–1459.
56. Holtschke, T., Löhler, J., Kanno, Y., Fehr, T., Giese, N., Rosenbauer, F., Lou, J., Knobeloch, K.P., Gabriele, L., Waring, J.F., et al. (1996). Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the ICSBP gene. Cell *87*, 307–317.
57. Fotiadis, D., Kanai, Y., and Palacín, M. (2013). The SLC3 and SLC7 families of amino acid transporters. Mol. Aspects Med. *34*, 139–158.
58. Yamanaka, R., Barlow, C., Lekstrom-Himes, J., Castilla, L.H., Liu, P.P., Eckhaus, M., Decker, T., Wynshaw-Boris, A., and Xanthopoulos, K.G. (1997). Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein epsilon-deficient mice. Proc. Natl. Acad. Sci. USA *94*, 13187–13192.
59. Halene, S., Gaines, P., Sun, H., Zibello, T., Lin, S., Khanna-Gupta, A., Williams, S.C., Perkins, A., Krause, D., and Berliner, N. (2010). C/EBPepsilon directs granulocytic-vs-monocytic lineage determination and confers chemotactic function via Hlx. Exp. Hematol. *38*, 90–103.
60. Francis, O.L., Payne, J.L., Su, R.J., and Payne, K.J. (2011). Regulator of myeloid differentiation and function: The secret life of Ikaros. World J. Biol. Chem. *2*, 119–125.
61. Domen, J., Cheshier, S.H., and Weissman, I.L. (2000). The role of apoptosis in the regulation of hematopoietic stem cells: Overexpression of Bcl-2 increases both their number and repopulation potential. J. Exp. Med. *191*, 253–264.

62. Elmore, S. (2007). Apoptosis: a review of programmed cell death. Toxicol. Pathol. *35*, 495–516.
63. Yoon, K.W., Byun, S., Kwon, E., Hwang, S.Y., Chu, K., Hiraki, M., Jo, S.H., Weins, A., Hakroush, S., Cebulla, A., et al. (2015). Control of signaling-mediated clearance of apoptotic cells by the tumor suppressor p53. Science *349*, 1261669.
64. Wang, L., Le Mercier, I., Putra, J., Chen, W., Liu, J., Schenk, A.D., Nowak, E.C., Suriawinata, A.A., Li, J., and Noelle, R.J. (2014). Disruption of the immune-checkpoint VISTA gene imparts a proinflammatory phenotype with predisposition to the development of autoimmunity. Proc. Natl. Acad. Sci. USA *111*, 14846–14851.
65. De Togni, P., Goellner, J., Ruddle, N.H., Streeter, P.R., Fick, A., Mariathasan, S., Smith, S.C., Carlson, R., Shornick, L.P., Strauss-Schoenberger, J., et al. (1994). Abnormal development of peripheral lymphoid organs in mice deficient in lymphotoxin. Science *264*, 703–707.
66. Chiquet-Ehrismann, R., and Tucker, R.P. (2011). Tenascins and the importance of adhesion modulation. Cold Spring Harb. Perspect. Biol. *3*, 3.
67. Yoshida, H., Okada, T., Haze, K., Yanagi, H., Yura, T., Negishi, M., and Mori, K. (2001). Endoplasmic reticulum stress-induced formation of transcription factor complex ERSF including NF-Y (CBF) and activating transcription factors 6alpha and 6beta that activates the mammalian unfolded protein response. Mol. Cell. Biol. *21*, 1239–1248.
68. Yamamoto, M., Ma, J.S., Mueller, C., Kamiyama, N., Saiga, H., Kubo, E., Kimura, T., Okamoto, T., Okuyama, M., Kayama, H., et al. (2011). ATF6beta is a host cellular target of the *Toxoplasma gondii* virulence factor ROP18. J. Exp. Med. *208*, 1533–1546.
69. Conduit, P.T., Wainman, A., and Raff, J.W. (2015). Centrosome function and assembly in animal cells. Nat. Rev. Mol. Cell Biol. *16*, 611–624.
70. Zhang, Q., Zhu, H., Xu, X., Li, L., Tan, H., and Cai, X. (2015). Inactivated Sendai virus induces apoptosis and autophagy via the PI3K/Akt/mTOR/p70S6K pathway in human non-small cell lung cancer cells. Biochem. Biophys. Res. Commun. *465*, 64–70.
71. Utsugi, M., Dobashi, K., Ono, A., Ishizuka, T., Matsuzaki, S., Hisada, T., Shimizu, Y., Kawata, T., Aoki, H., Kamide, Y., and Mori, M. (2009). PI3K p110beta positively regulates lipopolysaccharide-induced IL-12 production in human macrophages and dendritic cells and JNK1 plays a novel role. J. Immunol. *182*, 5225–5231.
72. Dehghan, A., Dupuis, J., Barbalic, M., Bis, J.C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski, H., Kettunen, J., Henneman, P., et al. (2011). Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation *123*, 731–738.
73. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. *45*, 1238–1243.
74. Yanaba, K., Bouaziz, J.D., Matsushita, T., Magro, C.M., St Clair, E.W., and Tedder, T.F. (2008). B-lymphocyte contributions to human autoimmune disease. Immunol. Rev. *223*, 284–299.
75. Orrù, V., Steri, M., Sole, G., Sidore, C., Virdis, F., Dei, M., Lai, S., Zoledziewska, M., Busonero, F., Mulas, A., et al. (2013). Genetic variants regulating immune cell levels in health and disease. Cell *155*, 242–256.
76. Sancho, D., Gómez, M., Viedma, F., Esplugues, E., Gordón-Alonso, M., García-López, M.A., de la Fuente, H., Martínez-A, C., Lauzurica, P., and Sánchez-Madrid, F. (2003). CD69 downregulates autoimmune reactivity through active transforming growth factor-beta production in collagen-induced arthritis. J. Clin. Invest. *112*, 872–882.
77. Hu, X., Kim, H., Stahl, E., Plenge, R., Daly, M., and Raychaudhuri, S. (2011). Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. Am. J. Hum. Genet. *89*, 496–506.
78. Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I., et al. (2014). Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science *344*, 519–523.
79. Malik, M., Simpson, J.F., Parikh, I., Wilfred, B.R., Fardo, D.W., Nelson, P.T., and Estus, S. (2013). CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing. J. Neurosci. *33*, 13320–13325.
80. Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., Coin, L., de Silva, R., Cookson, M.R., et al.; UK Brain Expression Consortium; North American Brain Expression Consortium (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat. Neurosci. *17*, 1418–1428.
81. Hambleton, S., Salem, S., Bustamante, J., Bigley, V., Boisson-Dupuis, S., Azevedo, J., Fortin, A., Haniffa, M., Ceron-Gutierrez, L., Bacon, C.M., et al. (2011). IRF8 mutations and human dendritic-cell immunodeficiency. N. Engl. J. Med. *365*, 127–138.
82. Hancock, A.M., Witonsky, D.B., Alkorta-Aranburu, G., Beall, C.M., Gebremedhin, A., Sukernik, R., Utermann, G., Pritchard, J.K., Coop, G., and Di Rienzo, A. (2011). Adaptations to climate-mediated selective pressures in humans. PLoS Genet. *7*, e1001375.
83. Cserti, C.M., and Dzik, W.H. (2007). The ABO blood group system and *Plasmodium falciparum* malaria. Blood *110*, 2250–2258.
84. Rijneveld, A.W., Levi, M., Florquin, S., Speelman, P., Carmeliet, P., and van Der Poll, T. (2002). Urokinase receptor is necessary for adequate host defense against pneumococcal pneumonia. J. Immunol. *168*, 3507–3511.
85. Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E., et al. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell *7*, 532–544.
86. Dowell, J.A., Korth-Bradley, J., Liu, H., King, S.P., and Berger, M.S. (2001). Pharmacokinetics of gemtuzumab ozogamicin, an antibody-targeted chemotherapy agent for the treatment of patients with acute myeloid leukemia in first relapse. J. Clin. Pharmacol. *41*, 1206–1214.
87. de Paulis, A., Montuori, N., Prevete, N., Fiorentino, I., Rossi, F.W., Visconte, V., Rossi, G., Marone, G., and Ragno, P. (2004). Urokinase induces basophil chemotaxis through a urokinase receptor epitope that is an endogenous ligand for formyl peptide receptor-like 1 and -like 2. J. Immunol. *173*, 5739–5748.
88. Selleri, C., Montuori, N., Ricci, P., Visconte, V., Carriero, M.V., Sidenius, N., Serio, B., Blasi, F., Rotoli, B., Rossi, G., and Ragno, P. (2005). Involvement of the urokinase-type plasminogen activator receptor in hematopoietic stem cell mobilization. Blood *105*, 2198–2205.

89. Kwok, B., Hall, J.M., Witte, J.S., Xu, Y., Reddy, P., Lin, K., Flamholz, R., Dabbas, B., Yung, A., Al-Hafidh, J., et al. (2015). MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. Blood *126*, 2355–2361.
90. Cargo, C.A., Rowbotham, N., Evans, P.A., Barrans, S.L., Bowen, D.T., Crouch, S., and Jack, A.S. (2015). Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression. Blood *126*, 2362–2365.
91. Abdel-Wahab, O., Gao, J., Adli, M., Dey, A., Trimarchi, T., Chung, Y.R., Kuscu, C., Hricik, T., Ndiaye-Lobry, D., Lafave, L.M., et al. (2013). Deletion of Asxl1 results in myelodysplasia and severe developmental defects in vivo. J. Exp. Med. *210*, 2641–2659.
92. Dupuis, A., Gaub, M.P., Legrain, M., Drenou, B., Mauvieux, L., Lutz, P., Herbrecht, R., Chan, S., and Kastner, P. (2013). Biclonal and biallelic deletions occur in 20% of B-ALL cases with IKZF1 mutations. Leukemia *27*, 503–507.
93. Mullighan, C.G., Miller, C.B., Radtke, I., Phillips, L.A., Dalton, J., Ma, J., White, D., Hughes, T.P., Le Beau, M.M., Pui, C.H., et al. (2008). BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature *453*, 110–114.
94. Ouyang, W., Kolls, J.K., and Zheng, Y. (2008). The biological functions of T helper 17 cell effector cytokines in inflammation. Immunity *28*, 454–467.
95. Pistis, G., Okonkwo, S.U., Traglia, M., Sala, C., Shin, S.Y., Masciullo, C., Buetti, I., Massacane, R., Mangino, M., Thein, S.L., et al.; CHARGE Consortium Hematology Working (2013). Genome wide association analysis of a founder population identified TAF3 as a gene for MCHC in humans. PLoS ONE *8*, e69206.
96. Migliorini, G., Fiege, B., Hosking, F.J., Ma, Y., Kumar, R., Sherborne, A.L., da Silva Filho, M.I., Vijayakrishnan, J., Koehler, R., Thomsen, H., et al. (2013). Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype. Blood *122*, 3298–3307.

The American Journal of Human Genetics, Volume *99*

Supplemental Data

Large-Scale Exome-wide Association Analysis Identifies Loci for White Blood Cell Traits and Pleiotropy with Immune-Mediated Diseases

Salman M. Tajuddin, Ursula M. Schick, John D. Eicher, Nathalie Chami, Ayush Giri, Jennifer A. Brody, W. David Hill, Tim Kacprowski, Jin Li, Leo-Pekka Lyytikäinen, Ani Manichaikul, Evelin Mihailov, Michelle L. O'Donoghue, Nathan Pankratz, Raha Pazoki, Linda M. Polfus, Albert Vernon Smith, Claudia Schurmann, Caterina Vacchi-Suzzi, Dawn M. Waterworth, Evangelos Evangelou, Lisa R. Yanek, Amber Burt, Ming-Huei Chen, Frank J.A. van Rooij, James S. Floyd, Andreas Greinacher, Tamara B. Harris, Heather M. Highland, Leslie A. Lange, Yongmei Liu, Reedik Mägi, Mike A. Nalls, Rasika A. Mathias, Deborah A. Nickerson, Kjell Nikus, John M. Starr, Jean-Claude Tardif, Ioanna Tzoulaki, Digna R. Velez Edwards, Lars Wallentin, Traci M. Bartz, Lewis C. Becker, Joshua C. Denny, Laura M. Raffield, John D. Rioux, Nele Friedrich, Myriam Fornage, He Gao, Joel N. Hirschhorn, David C.M. Liewald, Stephen S. Rich, Andre Uitterlinden, Lisa Bastarache, Diane M. Becker, Eric Boerwinkle, Simon de Denus, Erwin P. Bottinger, Caroline Hayward, Albert Hofman, Georg Homuth, Ethan Lange, Lenore J. Launer, Terho Lehtimäki, Yingchang Lu, Andres Metspalu, Chris J. O'Donnell, Rakale C. Quarells, Melissa Richard, Eric S. Torstenson, Kent D. Taylor, Anne-Claire Vergnaud, Alan B. Zonderman, David R. Crosslin, Ian J. Deary, Marcus Dörr, Paul Elliott, Michele K. Evans, Vilmundur Gudnason, Mika Kähönen, Bruce M. Psaty, Jerome I. Rotter, Andrew J. Slater, Abbas Dehghan, Harvey D. White, Santhi K. Ganesh, Ruth J.F. Loos, Tõnu Esko, Nauder Faraday, James G. Wilson, Mary Cushman, Andrew D. Johnson, Todd L. Edwards, Neil A. Zakai, Guillaume Lettre, Alex P. Reiner, and Paul L. Auer

Supplemental Figure

Figure S1. Quantile-quantile plots of p-values of white blood cell traits. Results of the three sets of discovery meta-analyses in combined all ancestries (ALL), European ancestry (EA), and African ancestry (AA) individuals are shown here. The combined all ancestry samples include Hispanic Americans, East Asians, and South Asians in addition to EAs and AAs.

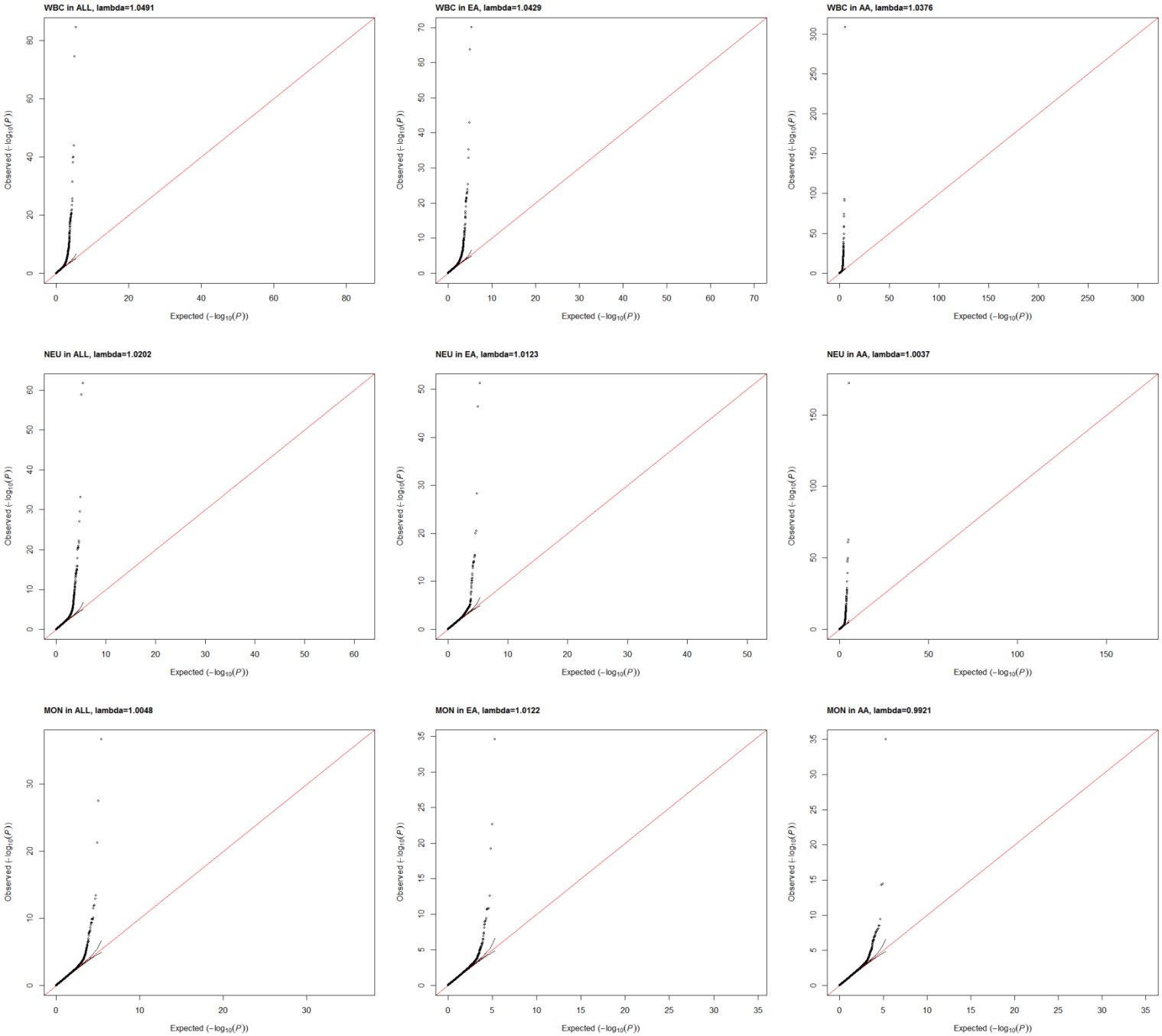


Figure S1. Continued

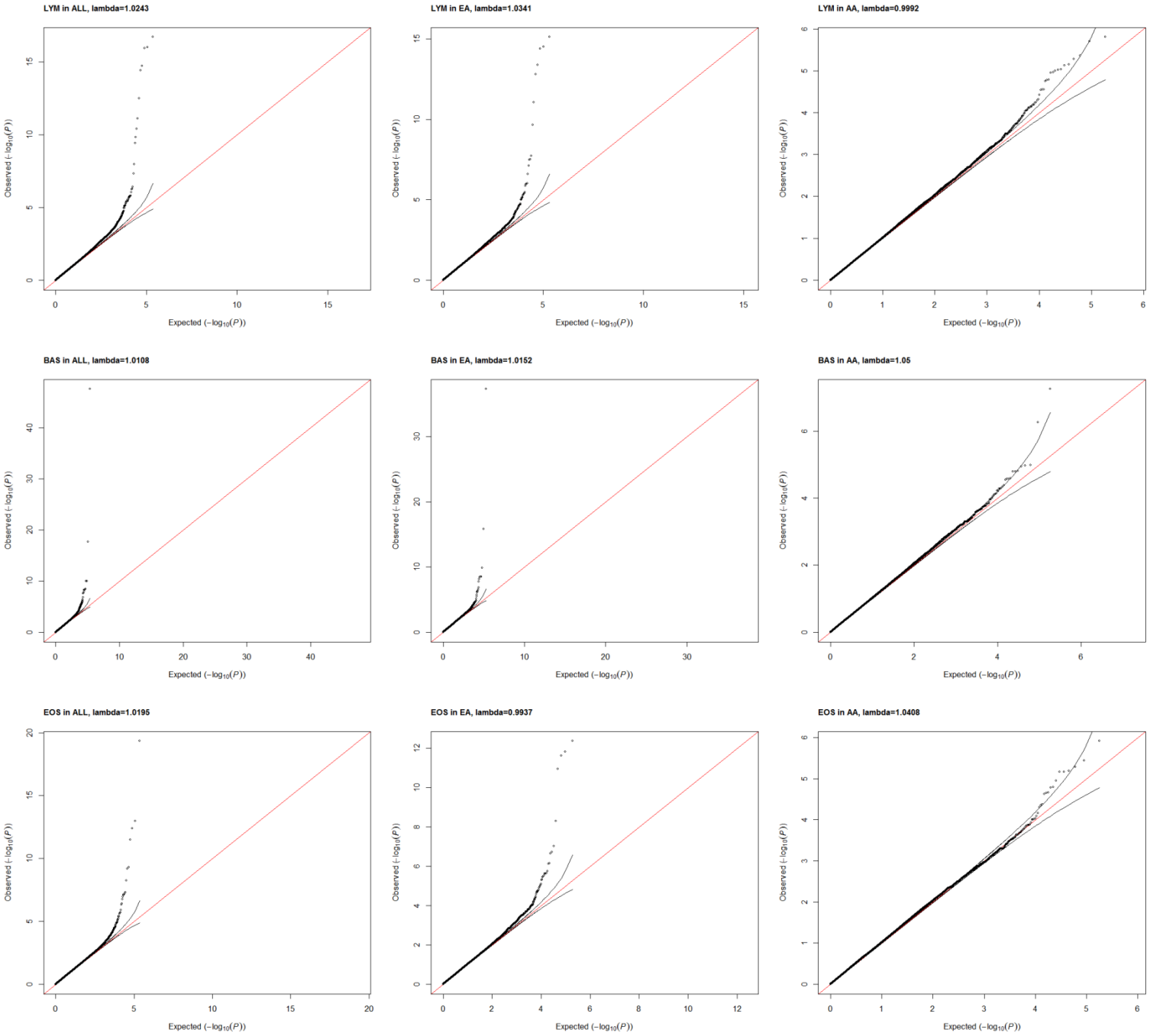
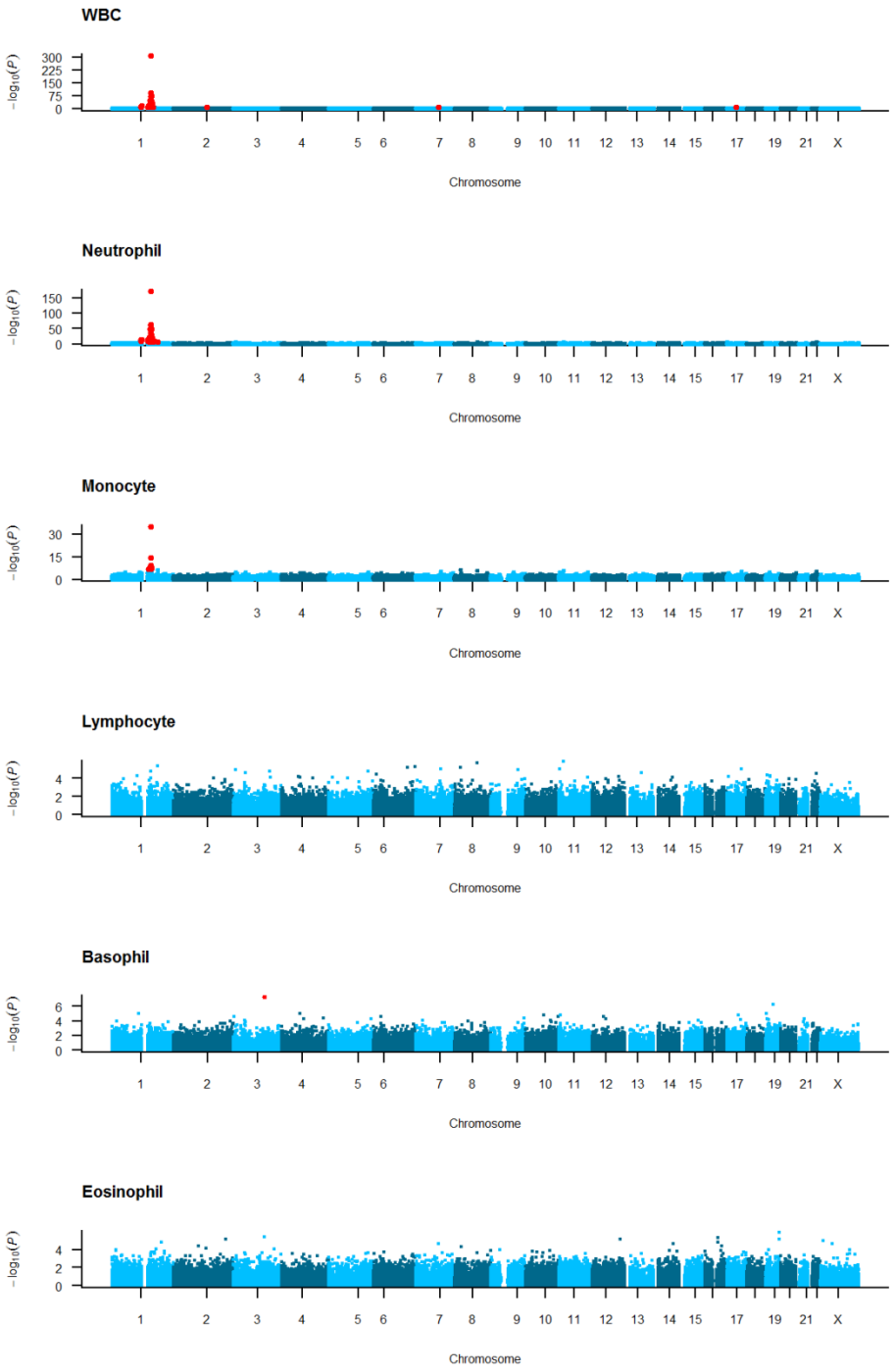


Figure S2. Manhattan plots of discovery p-values of white blood cell traits in African Americans



Supplemental Tables

See Table S1 and Table S2 in the accompanying Excel file.

Table S3. Sample size across ethnicities for white blood cell exome-wide association analysis

Study	Total WBC					Neutrophil					Monocyte			Lymphocyte			Basophil			Eosinophil		
	Ethnicity					Ethnicity					Ethnicity			Ethnicity			Ethnicity			Ethnicity		
	EA	AA	HA	EAS	SA	EA	AA	HA	EAS	SA	EA	AA	HA	EA	AA	HA	EA	AA	HA/HL	EA	AA	HA
Discovery																						
AGES	2,955					2,954					2,954	368		2,954						2,954		
ARIC	10,345	2,821				6,792	368				6,792			6,792	368		6,792	368		6,792	368	
BIOME	1,043	2,328	3,319			636	1,981	2,688			10,209	1,981	2,688	636	1,981	2,688	636	1,981	2,688	636	1,981	2,688
BIOVU - I	21,276	1,921				10,208	909				636	910		10,218	911		14,903	1,357				
BIOVU - II		1,020					385					384			385			393				
CARDIA	2,110	1,946				2,110	1,942				2,095	1,928		2,110	1,942		1,637	1,350		1,955	1,762	
CHS	4,197	776																				
ECGUT	974					973					973			973			962			968		
FHS	5,610					2,314					2,314			2,314			2,314			2,314		
SOLID TIMI-52	8,025	198	874	254	116	8,026	198	874	251	116												
STABILITY	8,490	118	500	714	348	8,487	118	500	714	347												
HABC	1,251					1,251					1,251			1,251			1,251			1,251		
HANDLS		834					834					834			834			813			832	
JHS		2,039					2,027					2,026			2,028			1,898			1,989	
LBC1921	475					473					473			473								
LBC1936	963					963					963			963								
MESA	1,266	750	843			1,122	673	763			1,123	675	764	1,121	675	762	1,123	673	765	1,121	671	762
MHIBB	7,911					6,978					6,978			6,978			6,978			6,978		
REGARDS		5,030																				
RS	2,794													2,758								
SHIP	7,251					4,076					4,076			4,076			4,075			4,077		
WHI	21,660	3,469				3,488	684				3,488	684		3,488	684		3,467	676		3,471	679	
Total per ancestry	108,596	23,250	5,536	968	464	60,851	10,119	4,825	965	463	44,325	9,790	3,452	47,105	9,808	3,450	44,138	9,509	3,453	32,517	8,282	3,450
Total per trait					**138,814**					**77,223**			**57,567**			**60,363**			**57,100**			**44,249**
Replication																						
AIRWAVE	14,887					14,887					14,887			14,887			14,887			14,887		
BIOVU-Replication (EA)	1,987					888					888			888			905					
FINCAVAS	911					397					397			397			397			440		
GeneSTAR	1,023					894					894			937								
Total per ancestry	**18,808**					**17,066**					**17,066**			**17,109**			**16,189**			**15,327**		
Total per trait					18,808					17,066			17,066			17,109			16,189			15,327
Grand Total					**157,622**					**94,289**			**74,633**			**77,472**			**73,289**			**59,576**

Abbreviations: EA, European Ancestry; AA, African Ancestry; HA, Hispanic American; EAS, East Asian; SA, South Asian.

Table S4. Mean and standard deviations of white blood cell traits in participating cohorts of the Blood-Cell Consortium

Study	Age	% Female	Total WBC	Neutrophil	Monocyte	Lymphocyte	Basophil	Eosinophil
AIRWAVE	40.87 (9.03)	37.3	6.52 (1.71)	3.94 (1.34)	0.4 (0.18)	1.8 (0.53)	0.06 (0.04)	0.2 (0.12)
n	14887	5550	14887	14886	14887	14886	14865	14866
AGES	76.4 (5.5)	57.9	6.05 (1.81)	3.55 (1.30)	0.541 (0.0178)	1.74 (0.972)	0.0286 (0.0256)	0.207 (0.145)
n	2953	1711	2955	2954	2954	2954	2954	2954
ARIC-AA	56.5 (5.8)	62.8	5.6 (1.9)	3.1 (1.4)	0.3 (0.1)	2.0 (0.6)	0.1 (0.0)	0.1 (0.1)
n	2832	1779	2832	369	369	369	160	355
ARIC-EA	57.3 (5.7)	53	6.2 (2.0)	3.8 (1.4)	0.4 (0.2)	1.8 (0.7)	0.1 (0.0)	0.2 (0.1)
n	10347	5489	10346	6792	6792	6792	2351	3893
BioME-AA	57 (14)	38.6	6.6 (2.2)	3.8 (1.8)	0.5 (0.2)	2.0 (0.8)	0.0 (0.0)	0.2 (0.1)
n	2653	1024	2329	1966	1966	1966	1966	1966
BioME-EA	68 (11)	38.9	7.2 (2.9)	4.7 (2.4)	0.6 (0.2)	1.7 (1.6)	0.0 (0.0)	0.2 (0.1)
n	1090	424	1046	634	634	634	634	634
BioME-HA	61 (15)	62	7.4 (2.3)	4.6 (1.9)	0.6 (0.2)	2.0 (0.8)	0.0 (0.0)	0.2 (0.2)
n	3640	2258	3319	2660	2660	2660	2660	2660
BioVU-I-EA	56.42 (16.82)	54.8	7.87 (5.34)	5.03 (3.16)	0.64 (0.54)	1.88 (2.76)	0.0456 (0.0433)	NA
N	21276	11654	21276	10208	10209	10218	14903	NA
BioVU-I-AA	49.58 (18.37)	61.1	7.30 (3.45)	4.56 (0.04)	0.59 (0.32)	2.06 (3.44)	0.0337 (0.0341)	NA
n	1921	1174	1921	909	910	911	1357	NA
BioVU-II-EA	50.93 (15.22)	100	7.75 (2.58)	4.93 (2.58)	0.58 (0.27)	1.9 (0.81)	0.036 (0.026)	NA
n	2298	2298	1987	888	888	888	928	NA
BioVU-II-AA	39.40 (13.60)	100	7.63 (3.17)	4.58 (3.14)	0.56 (0.33)	2.53 (5.91)	0.0346 (0.0497)	NA
n	1375	1375	1020	385	384	385	393	NA
CARDIA-EA	25.41 (3.36)	52.7	6.25 (1.78)	3.50 (1.40)	0.32 (0.18)	2.13 (0.65)	0.04 (0.05)	0.17 (0.14)
n	2110	1113	2110	2110	2095	2110	1637	1955
CARDIA-AA	24.23 (3.77)	56.3	5.90 (1.95)	3.12 (1.49)	0.31 (0.17)	2.24 (0.83)	0.05 (0.04)	0.16 (0.13)
n	1946	1095	1946	1942	1928	1942	1348	1760
CHS-EA	72.8 (5.6)	56.3	6.4 (2.0)	NA	NA	NA	NA	NA
n	4197	2361	4197	NA	NA	NA	NA	NA
CHS-AA	72.8 (5.6)	62.2	5.9 (1.9)	NA	NA	NA	NA	NA
n	776	483	776	NA	NA	NA	NA	NA
EGCUT	47.08 (16.72)	46.9	6.36 (1.7842)	3.70 (1.4080)	0.51 (0.1773)	1.96 (0.6362)	0.0299 (0.0194)	0.154 (0.1217)
n	974	457	974	973	973	973	974	973
FHS	42.6 (13.9)	53.1	6.1 (1.6)	3.6 (1.2)	0.5 (0.1)	1.7 (0.5)	0.0 (0.0)	0.2 (0.1)
n	5790	3077	5610	2315	2315	2315	2315	2315
FINCAVAS	53.3 (13.9)	41.9	7.3 (2.5)	4.1 (2.2)	0.5 (0.2)	2.2 (1.1)	0.04 (0.03)	0.23 (0.2)
n	913	383	912	396	396	396	396	436
GeneSTAR-EA	42.47 (12.67)	52.20	6.64 (2.00)	4.21 (1.55)	0.38 (0.18)	2.09 (0.63)	NA	NA
n	1023	534	1023	894	894	937	NA	NA
GeneSTAR-AA	42.40 (12.34)	61.99	6.01 (1.87)	3.47 (1.47)	0.3 (0.16)	2.23 (0.7)	NA	NA
n	613	380	613	543	543	562	NA	NA
HABC-EA	75.72 (2.83)	52.5	6.593 (3.69)	3.837 (1.44)	0.564 (0.20)	1.93 (3.20)	0.0636 (0.04)	0.19 (0.17)
n	1251	657	1251	1251	1251	1251	1251	1251
HANDLS	48.13 (8.90)	54.7	6.23 (2.06)	3.55 (1.66)	0.40 (0.16)	2.08 (0.72)	0.03 (0.02)	0.16 (0.13)
n	870	476	834	834	834	834	813	832
JHS	53.0 (13.0)	37.5	5.59 (1.87)	3.09 (1.48)	0.39 (0.14)	1.93 (0.66)	0.03 (0.02)	0.14 (0.13)
n	2258	846	2044	2032	2030	2032	1903	2000
LBC1936	69.6 (0.8)	49.6	7.0 (1.9)	4.4 (1.5)	0.5 (0.2)	1.8 (0.8)	NA	NA
n	964	478	963	963	963	963	NA	NA
LBC1921	79.1 (0.6)	58.3	7.1 (1.9)	4.6 (1.4)	0.5 (0.2)	1.7 (1.1)	NA	NA
n	475	277	475	473	473	473	NA	NA
MESA-EA	69.49 (9.55)	50.1	6.139 (1.706)	3.79 (1.299)	0.483 (0.1842)	1.666 (0.5854)	0.028 (0.0445)	0.171 (0.1275)
n	1267	635	1266	1193	1194	1193	1179	1186
MESA-AA	69.54 (9.10)	53.9	5.779 (1.93)	3.205 (1.4436)	0.428 (0.166)	1.883 (0.6357)	0.018 (0.0341)	0.162 (0.1284)
n	750	404	750	669	670	670	653	667
MESA-HA	68.93 (9.43)	52	6.185 (1.6204)	3.68647 (1.3285)	0.433 (0.1618)	1.789 (0.5582)	0.020 (0.0385)	0.1721(0.1322)
n	845	439	843	623	622	622	609	613
MESA-EAS	65.12 (10.01)	47.1	5.314 (1.3629)	3.0834 (1.0671)	0.290 (0.108)	1.640 (0.5039)	0.030 (0.0311)	0.122 (0.1224)
n	119	56	119	94	94	94	82	86
MHIBB	63.9 (10.9)	36.6	7.3 (2.5)	4.6 (2.2)	0.6 (0.2)	1.8 (1.0)	0.0 (0.1)	0.2 (0.1)
n	7911	2899	7911	7911	7911	7911	7911	7911
REGARDS	63.5 (9.2)	68.0	5.6 (2.0)	NA	NA	NA	NA	NA
n	5039	3427	5030	NA	NA	NA	NA	NA
RS	69.8 (0.7)	53.6	7(2)	NA	NA	3 (1)	NA	NA

n	3016	1618	2794	NA	NA	2758	NA	NA
SHIP	49.19 (16.09)	51	6.74 (2.02)	NA	NA	NA	NA	NA
n	3164	1613	3159	NA	NA	NA	NA	NA
SHIP-Trend	52.00 (15.40)	51.3	6.15 (1.94)	3.64 (1.41)	0.53 (0.18)	1.75 (0.59)	0.03 (0.02)	0.16 (0.12)
n	4245	2179	4238	4222	4222	4222	4221	4223
SOLID-TIMI 52-AA	60.50 (10.13)	33.8	6.909 (2.137)	4.3402 (1.773)	NA	NA	NA	NA
n	201	68	200	200	NA	NA	NA	NA
SOLID-TIMI 52-EA	64.55 (9.25)	25.8	7.5738 (2.22)	5.0038 (1.81)	NA	NA	NA	NA
n	8111	2092	8035	8036	NA	NA	NA	NA
SOLID-TIMI 52-EAS	63.59 (10.4)	19.6	6.8109 (2.005)	4.2585 (1.65)	NA	NA	NA	NA
n	255	50	255	252	NA	NA	NA	NA
SOLID-TIMI 52-SA	58.28 (9.67)	15.4	8.1573 (2.396)	5.6544 (2.074)	NA	NA	NA	NA
n	117	18	116	116	NA	NA	NA	NA
SOLID-TIMI 52-HA	63.29 (9.74)	28.3	7.4106 (2.171)	4.8 (1.792)	NA	NA	NA	NA
n	893	253	875	875	NA	NA	NA	NA
STABILITY-AA	62.86 (10.22)	25.4	5.6873 (1.67)	3.3362 (1.379)	NA	NA	NA	NA
n	118	30	118	118	NA	NA	NA	NA
STABILITY-EA	65.91 (8.98)	17.6	6.8245 (1.839)	4.3007 (1.427)	NA	NA	NA	NA
n	8499	1499	8490	8487	NA	NA	NA	NA
STABILITY-EAS	64.25 (8.72)	22.3	6.7628 (1.87)	4.1576 (1.466)	NA	NA	NA	NA
n	714	159	714	714	NA	NA	NA	NA
STABILITY-SA	58.35 (10.84)	13	7.5481 (1.977)	4.8583 (1.697)	NA	NA	NA	NA
n	354	46	348	347	NA	NA	NA	NA
STABILITY-HL	66.34 (8.41)	17.2	6.6986 (1.768)	4.2348 (1.444)	NA	NA	NA	NA
n	500	86	500	500	NA	NA	NA	NA
WHI-EA	66.2 (6.7)	100	6.1 (1.7)	3.9 (1.4)	0.6 (0.2)	1.8 (1.0)	0.0 (0.0)	0.2 (0.1)
n	21867	21867	21841	3669	3669	3669	3669	3669
WHI-AA	65.0 (6.6)	100	5.6 (1.7)	3.2 (1.5)	0.5 (0.2)	1.9 (1.0)	0.0 (0.0)	0.2 (0.2)
n	3501	3501	3497	712	712	712	713	712
Total	134,770		133,273	78,123	58,450	61,248	56,263	42,615

Abbreviations: EA, European Ancestry; AA, African Ancestry; HA, Hispanic American; EAS, East Asian; SA, South Asian.

Table S5. Blood cell specific expression quantitative trait loci (eQTL) datasets used for lookups of white blood cell trait variants identified in the present study

Datasets	Reference [PMID]
Blood cell related eQTL studies included fresh lymphocytes	17873875
Fresh leukocytes	19966804
Leukocyte samples in individuals with Celiac disease	19128478
Whole blood samples	18344981, 21829388, 22692066, 23818875, 23359819, 23880221, 24013639, 23157493, 23715323, 24092820, 24314549, 24956270, 24592274, 24728292, 24740359, 25609184, 22563384, 25474530, 25816334, 25578447
Lymphoblastoid cell lines derived from asthmatic children	17873877, 23345460
Lymphoblastoid cell lines derived from 3 HapMap populations	17873874
Lymphoblastoid cell lines derived from HapMap CEU populations	18193047
Lymphoblastoid cell lines from population samples	19644074, 22286170, 22941192, 23755361, 23995691, 25010687, 25951796
Neutrophils	26151758, 26259071
CD19+ B cells	22446964
Primary phytohaemagglutinin (PHA)-stimulated T cells	19644074, 23755361
CD4+ T cells	20833654
Peripheral blood monocytes	19222302,20502693,22446964, 23300628, 25951796, 26019233
CD14+ monocytes before and after stimulation with lipopolysaccharide (LPS) or interferon-gamma	24604202
CD11+ dendritic cells before and after Mycobacterium tuberculosis infection	22233810
Dendritic cells before or after stimulation with lipopolysaccharide (LPS), influenza or interferon-beta	24604203
Micro-RNA QTLs	21691150, 26020509
DNase-I QTLs	22307276
Histone acetylation QTLs	25799442
Ribosomal occupancy QTLs	25657249
Splicing QTLs	25685889
Micro-RNA QTLs	25791433
Splicing QTLs	25685889
ScanDB eQTL data[a]	http://www.scandb.org/newinterface/about.html
Genotype-Tissue Expression (GTEx) whole blood eQTL[b]	23715323, www.gtexportal.org
eQTL databases at University of Chicago	http://eqtl.uchicago.edu/Home.html
Long non-coding RNAs in monocytes	25025429

[a]ScanDB cis-eQTLs were limited to those with P<1.0E-6 and trans-eQTLs with P<5.0E-8.

[b]GTEx Analysis V4 for 13 tissues were downloaded from the GTEx Portal (www.gtexportal.org) and results were filtered using the following criteria: Splicing QTL (sQTL) results generated with sQTLseeker with false discovery rate P≤0.05 were retained. For all gene-level eQTLs, if at least 1 SNP passed the tissue-specific empirical threshold in GTEx, the best SNP for that eQTL was always retained. All gene-level eQTL SNPs with P<1.67E-11 were also retained, reflecting a global threshold correction of P=0.05/(30,000 genes X 1,000,000 tests).

See Table S6 in the accombanying Excel file.

Table S7. Independent single nucleotide variants (N=28) associated with white blood cell traits in European only and combined All ancestries meta-analysis

											Discovery				Replication				
Trait (population)	**dbSNPID**	**Chr**	**Pos**	**Alt**	**Ref**	**EAF**	**Gene**	**Annotation**	**AA Substitution**	**CADD Score**	**N**	**Beta**	**SE**	**P**	**N**	**EAF**	**Beta**	**SE**	**P**
WBC (EA)	rs4925663	1	247,614,617	T	C	0.39	*OR2B11*	Missense	p.Gly223Asp	24.10	107,345	0.0251	0.0046	4.55E-08	17,897	0.4	0.01693	0.0112	0.132
WBC (All)	rs4925663	1	247,614,617	T	C	0.37	*OR2B11*	Missense	p.Gly223Asp	24.10	131,513	0.026	0.0042	5.85E-10	17,897	0.4	0.01693	0.0112	0.132
WBC (All)	rs1260326	2	27,730,940	C	T	0.62	*GCKR*	Missense	p.Leu446Pro	0.11	132,764	-0.0299	0.0044	1.44E-11	17,897	0.6	-0.0437	0.012	0.00027
WBC (EA)	rs1260326	2	27,730,940	C	T	0.58	*GCKR*	Missense	p.Leu446Pro	0.11	108,596	-0.0298	0.0048	4.01E-10	17,897	0.6	-0.0437	0.012	0.00027
WBC (All)	rs2276853	3	47,282,303	A	G	0.58	*KIF9*	Missense	p.Arg573Trp	32.00	132,764	0.023	0.0042	3.65E-08	17,897	0.6	0.02516	0.0116	0.03
WBC (All)	rs9374080	6	109,616,420	C	T	0.43	*CCDC162P*	Intronic		3.79	132,764	0.0227	0.0041	4.01E-08	17,897	0.46	0.02512	0.0113	0.0255
LYM (All)	rs2229094	6	31,540,556	C	T	0.26	*LTA*	Missense	p.Cys13Arg	0.03	59,978	0.0451	0.0071	1.55E-10	16,711	0.25	0.07796	0.0175	8.54E-06
LYM (EA)	rs2229094	6	31,540,556	C	T	0.26	*LTA*	Missense	p.Cys13Arg	0.03	47,105	0.0456	0.0081	1.89E-08	16,711	0.25	0.07796	0.0175	8.54E-06
NEU (All)	rs185819[a]	6	32,050,067	C	T	0.52	*TNXB*	Missense	p.His1161Arg	0.00	76,838	0.0308	0.0058	9.65E-08	16,669	0.47	0.02635	0.0158	0.0948
WBC (All)	rs185819	6	32,050,067	C	T	0.51	*TNXB*	Missense	p.His1161Arg	0.00	132,764	0.0307	0.0049	4.02E-10	17,897	0.47	0.03395	0.0149	0.0224
WBC (All)	rs1050331	7	44,808,091	G	T	0.48	*ZMIZ2*	3'UTR		5.30	123,297	0.0235	0.0041	1.25E-08	15,910	0.48	0.01149	0.0117	0.325
MON (All)	rs4917014	7	50,305,863	G	T	0.28	*C7orf72-IKZF1*	Intergenic		1.91	57,183	-0.038	0.0068	1.97E-08	16,669	0.32	-0.0478	0.0122	8.92E-05
MON (EA)	rs10107630	8	130,603,635	T	C	0.57	*CCDC26*	Intronic		5.39	44,325	-0.0683	0.0069	2.45E-23	15,781	0.57	-0.1069	0.0118	1.06E-19
MON (All)	rs10107630	8	130,603,635	T	C	0.56	*CCDC26*	Intronic		5.39	57,183	-0.0663	0.006	3.11E-28	15,781	0.57	-0.1069	0.0118	1.06E-19
WBC (All)	rs1982151	9	86,617,265	G	A	0.71	*RMI1*	Missense	p.Asn455Ser	0.00	132,764	0.0255	0.0044	6.97E-09	17,897	0.74	0.02095	0.0123	0.0896
WBC (EA)	rs4409764	10	101,284,237	G	T	0.51	*GOT1-LINC01475*	Intergenic		7.60	108,596	-0.0255	0.0046	3.34E-08	17,897	0.53	-4.00E-04	0.0115	0.974
MON (All)	rs6584283	10	101,291,593	C	T	0.55	*LINC01475*	Intronic ncRNA		-	57,183	-0.035	0.0061	9.55E-09	16,669	0.53	-0.0121	0.0121	0.319
WBC (EA)	rs1935	10	64,927,823	G	C	0.49	*JMJD1C*	Missense	p.Glu2353Asp	16.64	108,596	-0.0261	0.0047	2.46E-08	17,897	0.46	-0.0271	0.0117	0.0206
NEU (All)	rs3747869	10	73,520,632	C	A	0.9	*C10orf54 (DD1α)*	Missense	p.Asp187Glu	10.88	76,838	0.0494	0.0087	1.58E-08	16,669	0.9	0.07326	0.019	0.00012
WBC (EA)	rs3747869	10	73,520,632	C	A	0.9	*C10orf54 (DD1α)*	Missense	p.Asp187Glu	10.88	108,596	0.0398	0.0073	4.26E-08	17,897	0.9	0.08289	0.0184	6.40E-06
NEU (EA)	rs3747869	10	73,520,632	C	A	0.9	*C10orf54 (DD1α)*	Missense	p.Asp187Glu	10.88	60,851	0.0533	0.0095	2.11E-08	16,669	0.9	0.07326	0.019	0.00012
WBC (All)	rs3747869	10	73,520,632	C	A	0.91	*C10orf54 (DD1α)*	Missense	p.Asp187Glu	10.88	132,764	0.0381	0.0068	2.31E-08	17,897	0.9	0.08289	0.0184	6.40E-06
NEU (EA)	rs144349650	11	55,432,976	G	C	0.00069	*OR4C6*	Missense	p.Leu112Val	13.09	60,851	0.7176	0.1096	5.95E-11	16,669	Monomorphic	NA	NA	NA
NEU (All)	rs144349650	11	55,432,976	G	C	0.00071	*OR4C6*	Missense	p.Leu112Val	13.09	76,838	0.6648	0.0962	4.92E-12	16,669	Monomorphic	NA	NA	NA
WBC (EA)	rs144349650	11	55,432,976	G	C	0.00042	*OR4C6*	Missense	p.Leu112Val	13.09	108,596	0.7207	0.1073	1.87E-11	17,897	2.81E-05	-0.6178	0.9965	0.535
WBC (All)	rs144349650	11	55,432,976	G	C	0.00044	*OR4C6*	Missense	p.Leu112Val	13.09	132,764	0.6379	0.0942	1.26E-11	17,897	2.81E-05	-0.6178	0.9965	0.535
LYM (EA)	rs199694284	11	5,632,403	C	T	5.3E-05	*TRIM6*	Missense	p.Val258Ala	25.20	40,313	-2.3971	0.4458	7.56E-08	16,711	5.98E-05	-0.9067	0.6995	0.195
BAS (EA)	rs4430553	12	66,698,895	C	T	0.54	*HELB*	Missense	p.Leu191Pro	5.73	44,138	0.0404	0.0068	3.02E-09	15,770	0.53	0.02111	0.0116	0.0681
BAS (All)	rs4430553	12	66,698,895	C	T	0.56	*HELB*	Missense	p.Leu191Pro	5.73	56,707	0.0391	0.006	9.73E-11	15,770	0.53	0.02111	0.0116	0.0681
LYM (All)	rs4763879	12	9,910,164	A	G	0.32	*CD69*	Intronic		3.03	59,978	-0.0366	0.0064	1.08E-08	16,711	0.36	-0.0381	0.0119	0.00136
LYM (EA)	rs4763879	12	9,910,164	A	G	0.36	*CD69*	Intronic		3.03	47,105	-0.0379	0.0069	3.08E-08	16,711	0.36	-0.0381	0.0119	0.00136
MON (EA)	rs11625112	14	23,596,740	G	A	0.46	*SLC7A8*	Intronic		6.34	44,325	-0.0376	0.0068	3.82E-08	16,669	0.45	-0.031	0.0115	0.00704
WBC (EA)	rs2306331	15	51,217,361	C	T	0.46	*AP4E1*	Missense	p.Cys88Arg	6.49	100,571	-0.0249	0.0046	6.15E-08	15,910	0.47	-0.0174	0.0116	0.134
MON (EA)	rs11642873[a]	16	85,991,705	C	A	0.2	*IRF8-LINC01082*	Intergenic		0.75	44,325	0.0573	0.0085	1.41E-11	16,669	0.2	0.11251	0.0144	6.17E-15

MON (All)	rs11642873	16	85,991,705	C	A	0.17	*IRF8-LINC01082*	Intergenic		0.75	57,183	0.0561	0.0081	3.24E-12	16,669	0.2	0.11251	0.0144	6.17E-15
MON (All)	rs1292053	17	57,963,537	G	A	0.45	*TUBD1*	Missense	p.Met76Thr	5.53	57,183	-0.0359	0.006	2.55E-09	16,669	0.44	-0.0398	0.0116	0.00061
WBC (All)	rs1292053	17	57,963,537	G	A	0.45	*TUBD1*	Missense	p.Met76Thr	5.53	132,764	-0.0287	0.004	1.06E-12	17,897	0.44	-0.027	0.0111	0.0151
WBC (EA)	rs1292053	17	57,963,537	G	A	0.45	*TUBD1*	Missense	p.Met76Thr	5.53	108,596	-0.0304	0.0045	1.28E-11	17,897	0.44	-0.027	0.0111	0.0151
BAS (All)	rs736289	19	33,757,062	C	T	0.43	*SLC7A10-CEBPA*	Intergenic		0.41	56,707	0.0343	0.0061	2.34E-08	15,770	0.4	0.01229	0.0116	0.291
BAS (EA)	rs736289	19	33,757,062	C	T	0.39	*SLC7A10-CEBPA*	Intergenic		0.41	44,138	0.0402	0.0069	7.02E-09	15,770	0.4	0.01229	0.0116	0.291
NEU (All)	rs4760	19	44,153,100	G	A	0.14	*CD87 (PLAUR)*	Missense	p.Lue272Pro	24.20	71,415	-0.0471	0.0079	2.33E-09	16,669	0.15	-0.044	0.0159	0.00555
WBC (EA)	rs4760	19	44,153,100	G	A	0.16	*CD87 (PLAUR)*	Missense	p.Lue272Pro	24.20	85,685	-0.043	0.0068	2.51E-10	17,897	0.15	-0.0518	0.0153	0.00071
WBC (All)	rs4760	19	44,153,100	G	A	0.14	*CD87 (PLAUR)*	Missense	p.Lue272Pro	24.20	106,384	-0.0424	0.0065	8.63E-11	17,897	0.15	-0.0518	0.0153	0.00071
NEU (EA)	rs4760	19	44,153,100	G	A	0.16	*CD87 (PLAUR)*	Missense	p.Lue272Pro	24.20	56,112	-0.0467	0.0083	1.54E-08	16,669	0.15	-0.044	0.0159	0.00555
WBC (All)	rs3865444	19	51,727,962	A	C	0.28	*CD33*	Upstream		3.79	107,635	-0.034	0.005	6.25E-12	17,897	0.32	-0.0329	0.0118	0.00514
WBC (EA)	rs3865444	19	51,727,962	A	C	0.31	*CD33*	Upstream		3.79	86,936	-0.0367	0.0053	3.51E-12	17,897	0.32	-0.0329	0.0118	0.00514
LYM (All)	rs6136489	20	1,923,734	G	T	0.4	*SIRPA-PDYN*	Intergenic		2.52	59,978	-0.0406	0.0061	3.99E-11	16,711	0.33	-0.0096	0.0121	0.428
WBC (All)	rs6136489	20	1,923,734	G	T	0.39	*SIRPA-PDYN*	Intergenic		2.52	132,764	-0.0246	0.0042	4.03E-09	17,897	0.33	-0.0179	0.0116	0.122
LYM (EA)	rs6136489	20	1,923,734	G	T	0.34	*SIRPA-PDYN*	Intergenic		2.52	47,105	-0.0384	0.007	3.33E-08	16,711	0.33	-0.0096	0.0121	0.428
BAS (EA)	rs2295764	20	31,025,163	G	A	0.36	*ASXL1*	3'UTR		0.10	44,138	-0.0419	0.0071	3.28E-09	15,770	0.36	-0.031	0.012	0.00978
BAS (All)	rs2295764	20	31,025,163	G	A	0.33	*ASXL1*	3'UTR		0.10	56,707	-0.0367	0.00641	1.07E-08	15,770	0.36	-0.031	0.012	0.00978
WBC (All)	rs2836878	21	40,465,534	A	G	0.26	*ETS2-PSMG1*	Intergenic		1.82	132,764	-0.0246	0.0046	8.36E-08	17,897	0.26	-0.0264	0.0125	0.0344

Abbreviations: Chr, chromosome; Pos, basepair position; Alt, effect allele; Ref, reference allele; EAF, effect allele frequency; AA, amino acid; CADD, Combined annotation dependent depletion; EA, European ancestry, All, combined European, African, Hispanic American, East Asian and South Asian ancestries; WBC, white blood cell; NEU, neutrophil; MON, monocyte; LYM, lymphocyte; BAS, basophil.

[a]Secondary signal identified through conditional analysis.

Table S8. Association between 16 replicated loci and white blood cell subtypes in the discovery meta-analysis. Primary associations are shown in bold.

dbSNPID	Chr	Pos	Gene	Alt/Ref	Trait (population)	N	EAF	Beta	SE	P
rs1260326	2	27,730,940	*GCKR*	C/T	**WBC (EA)**	**108,596**	**0.58**	**-0.030**	**0.005**	**4.01E-10**
					WBC (All)	132,764	0.62	-0.030	0.004	1.44E-11
					WBC (AA)	17,200	0.85	-0.039	0.016	1.44E-02
					NEU (EA)	60,851	0.58	-0.020	0.006	9.81E-04
					NEU (All)	76,838	0.62	-0.019	0.006	6.46E-04
					LYM (EA)	47,105	0.59	-0.016	0.007	2.15E-02
					LYM (All)	59,978	0.64	-0.013	0.006	3.91E-02
					BAS (EA)	44,138	0.59	-0.022	0.007	2.17E-03
					BAS (All)	56,707	0.64	-0.021	0.006	1.04E-03
rs2276853	3	47,282,303	*KIF9*	A/G	**WBC (All)**	**132,764**	**0.58**	**0.023**	**0.004**	**3.65E-08**
					WBC (EA)	108,596	0.60	0.022	0.005	1.39E-06
					NEU (EA)	60,851	0.60	0.017	0.006	4.56E-03
					NEU (All)	76,838	0.59	0.018	0.005	8.24E-04
					LYM (EA)	47,105	0.60	0.015	0.007	2.70E-02
					LYM (All)	59,978	0.58	0.018	0.006	2.52E-03
					LYM (AA)	9,423	0.53	0.039	0.015	7.86E-03
rs2229094	6	31,540,556	*LTA*	C/T	**LYM (EA)**	**47,105**	**0.26**	**0.046**	**0.008**	**1.89E-08**
					LYM (All)	59,978	0.26	0.045	0.007	1.55E-10
					LYM (AA)	9,423	0.28	0.044	0.017	7.30E-03
					WBC (EA)	108,596	0.26	0.035	0.006	2.73E-09
					WBC (All)	132,764	0.26	0.031	0.005	5.40E-09
					NEU (All)	76,838	0.26	0.015	0.006	1.74E-02
					BAS (EA)	44,138	0.26	0.026	0.008	1.37E-03
					BAS (All)	56,707	0.27	0.020	0.007	5.59E-03
rs185819	6	32,050,067	*TNXB*	C/T	**WBC (All)**	**132,764**	**0.51**	**0.031**	**0.005**	**4.02E-10**
					WBC (EA)	108,596	0.52	0.029	0.006	2.62E-07
					WBC (AA)	17,200	0.47	0.035	0.012	4.38E-03
					NEU (EA)	60,851	0.52	0.025	0.007	1.33E-04
					NEU (All)	76,838	0.52	0.031	0.006	9.65E-08
					NEU (AA)	9,734	0.47	0.058	0.016	1.80E-04
					MON (EA)	44,325	0.51	0.020	0.008	1.14E-02
					MON (All)	57,183	0.51	0.021	0.007	1.89E-03
					LYM (EA)	47,105	0.51	0.015	0.007	3.93E-02
					LYM (All)	59,978	0.50	0.019	0.006	3.23E-03
					LYM (AA)	9,423	0.47	0.041	0.015	5.79E-03
					EOS (EA)	32,517	0.48	0.019	0.009	4.39E-02
					EOS (All)	44,249	0.48	0.016	0.008	3.58E-02
					BAS (EA)	44,138	0.51	0.020	0.007	6.87E-03
					BAS (All)	56,707	0.51	0.020	0.006	1.90E-03
rs9374080	6	109,616,420	*CCDC162P*	C/T	**WBC (All)**	**132,764**	**0.43**	**0.023**	**0.004**	**4.01E-08**
					WBC (EA)	108,596	0.44	0.021	0.005	2.67E-06
					WBC (AA)	17,200	0.36	0.024	0.011	3.74E-02
					NEU (EA)	60,851	0.44	0.015	0.006	1.08E-02
					NEU (All)	76,838	0.42	0.018	0.005	5.69E-04
					MON (All)	57,183	0.43	0.014	0.006	2.71E-02
					BAS (EA)	44,138	0.45	0.026	0.007	1.34E-04
					BAS (All)	56,707	0.43	0.024	0.006	9.85E-05

rs4917014	7	50,305,863	*C7orf72-IKZF1*	G/T	**MON (All)**	**57,183**	**0.28458**	**-0.038**	**0.00677**	**1.97E-08**
					MON (EA)	44,325	0.32243	-0.0336	0.00726	3.84E-06
					MON (AA)	9,406	0.08266	-0.0606	0.02683	2.40E-02
rs3747869	10	73,520,632	*C10orf54 (DD1α)*	C/A	**WBC (EA)**	**108,596**	**0.90**	**0.040**	**0.007**	**4.26E-08**
					WBC (All)	132,764	0.91	0.038	0.007	2.31E-08
					NEU (EA)	**60,851**	**0.90**	**0.053**	**0.010**	**2.11E-08**
					NEU (All)	76,838	0.90	0.049	0.009	1.58E-08
					MON (EA)	44,325	0.90	0.041	0.011	2.44E-04
					MON (All)	57,183	0.91	0.037	0.010	3.77E-04
					EOS (EA)	32,517	0.82	0.033	0.014	1.58E-02
					EOS (All)	44,249	0.85	0.025	0.012	4.23E-02
rs1935	10	64,927,823	*JMJD1C*	G/C	**WBC (EA)**	**108,596**	**0.49**	**-0.026**	**0.005**	**2.46E-08**
					WBC (All)	132,764	0.46	-0.025	0.004	6.57E-09
					NEU (EA)	60,851	0.49	-0.024	0.006	4.00E-05
					NEU (All)	76,838	0.46	-0.022	0.005	4.99E-05
					LYM (All)	59,978	0.45	-0.013	0.006	2.86E-02
					EOS (All)	44,249	0.41	-0.020	0.007	6.33E-03
rs4763879	12	9,910,164	*CD69*	A/G	**LYM (EA)**	**47,105**	**0.36**	**-0.038**	**0.007**	**3.08E-08**
					LYM (All)	59,978	0.32	-0.037	0.006	1.08E-08
					WBC (EA)	108,596	0.36	-0.014	0.005	2.63E-03
					WBC (All)	132,764	0.33	-0.013	0.004	2.94E-03
rs11625112	14	23,596,740	*SLC7A8*	G/A	**MON (EA)**	**44,325**	**0.46**	**-0.038**	**0.007**	**3.82E-08**
					MON (All)	57,183	0.46	-0.030	0.006	4.20E-07
					EOS (EA)	32,517	0.42	-0.021	0.008	1.17E-02
					EOS (All)	44,249	0.43	-0.021	0.007	2.77E-03
					BAS (EA)	44,138	0.46	-0.031	0.007	7.76E-06
					BAS (All)	56,707	0.46	-0.026	0.006	1.07E-05
rs11642873	16	85,991,705	*IRF8-LINC01082*	C/A	**MON (EA)**	**44,325**	**0.20**	**0.057**	**0.008**	**1.41E-11**
					MON (All)	57,183	0.17	0.056	0.008	3.24E-12
					EOS (EA)	32,517	0.18	0.036	0.010	5.94E-04
					EOS (All)	44,249	0.15	0.033	0.010	8.18E-04
					BAS (EA)	44,138	0.20	0.021	0.009	1.53E-02
					BAS (All)	56,707	0.17	0.016	0.008	4.73E-02
rs1292053	17	57,963,537	*TUBD1*	G/A	**WBC (EA)**	**108,596**	**0.45**	**-0.030**	**0.004**	**1.28E-11**
					WBC (All)	132,764	0.45	-0.029	0.004	1.06E-12
					WBC (AA)	17,200	0.48	-0.029	0.011	8.31E-03
					NEU (EA)	60,851	0.45	-0.023	0.006	1.04E-04
					NEU (All)	76,838	0.45	-0.020	0.005	1.28E-04
					MON (EA)	44,325	0.45	-0.032	0.007	2.25E-06
					MON (All)	**57,183**	**0.45**	**-0.036**	**0.006**	**2.55E-09**
					MON (AA)	9,406	0.48	-0.035	0.015	1.80E-02
					LYM (EA)	47,105	0.44	-0.016	0.007	1.46E-02
					LYM (All)	59,978	0.45	-0.013	0.006	3.05E-02
rs4760	19	44,153,100	*CD87 (PLAUR)*	G/A	**NEU (EA)**	**56,112**	**0.16**	**-0.047**	**0.008**	**1.54E-08**
					NEU (All)	71,415	0.14	-0.047	0.008	2.33E-09
					WBC (EA)	**85,685**	**0.16**	**-0.043**	**0.007**	**2.51E-10**
					WBC (All)	106,384	0.14	-0.042	0.007	8.63E-11
rs3865444	19	51,727,962	*CD33*	A/C	**WBC (EA)**	**86,936**	**0.31**	**-0.037**	**0.005**	**3.51E-12**
					WBC (All)	107,635	0.28	-0.034	0.005	6.25E-12

					NEU (EA)	57,363	0.31	-0.025	0.006	1.15E-04
					NEU (All)	72,666	0.28	-0.021	0.006	3.26E-04
					MON (EA)	40,837	0.31	-0.020	0.008	9.46E-03
					MON (All)	53,011	0.28	-0.016	0.007	2.51E-02
					EOS (EA)	29,046	0.28	-0.029	0.010	2.16E-03
					EOS (All)	40,099	0.25	-0.025	0.009	3.09E-03
rs2295764	20	31,025,163	*ASXL1*	G/A	**BAS (EA)**	**44,138**	**0.36**	**-0.042**	**0.007**	**3.28E-09**
					BAS (All)	56,707	0.33	-0.037	0.006	1.07E-08
					WBC (AA)	17,200	0.20	-0.029	0.013	3.23E-02
					MON (EA)	44,325	0.36	-0.032	0.007	5.64E-06
					MON (All)	57,183	0.33	-0.026	0.006	3.95E-05
					EOS (EA)	32,517	0.32	-0.027	0.009	2.14E-03
					EOS (All)	44,249	0.30	-0.025	0.008	9.81E-04
rs2836878	21	40,465,534	*ETS2-PSMG1*	A/G	**WBC (All)**	**132,764**	**0.26**	**-0.025**	**0.005**	**8.36E-08**
					WBC (EA)	108,596	0.27	-0.024	0.005	7.49E-07
					NEU (EA)	60,851	0.28	-0.025	0.006	1.08E-04
					NEU (All)	76,838	0.26	-0.025	0.006	3.38E-05
					MON (EA)	44,325	0.28	-0.026	0.008	7.28E-04
					MON (All)	57,183	0.25	-0.022	0.007	1.69E-03
					BAS (EA)	44,138	0.28	-0.019	0.008	1.29E-02
					BAS (All)	56,707	0.25	-0.018	0.007	9.80E-03

Abbreviations: Chr, chromosome; Pos, basepair position; Alt, effect allele; Ref, reference allele; EAF, effect allele frequency; EA, European ancestry; AA, African ancestry; All, combined European, African, Hispanic American, East Asian and South Asian ancestries; WBC, white blood cell; NEU, neutrophil; MON, monocyte; LYM, lymphocyte; EOS, eosinophils; BAS, basophils.

See Table S9 in the accompanying Excel file

Table S10. Genes associated with white blood cell and differential counts identified using gene-based association meta-analysis

					Number of variants	Discovery						Replication				
Trait	Population	Gene	Test	Chr		N	P	Beta	S.D.	MAF cutoff	cMAC	Test	N	P	Beta	S.D.
WBC	EA	*CXCR2*	SKAT	2	9	108,596	1.241E-14	-0.223	0.029	1	3,725	SKAT	2,898	0.0173	-0.406	0.169
WBC	All	*CXCR2*	SKAT	2	8	138,814	9.48E-15	-0.193	0.027	1	10,438					
WBC	EA	*CXCR2*	VT	2	9	108,596	7.239E-14	-0.223	0.029	0.012	3,725	VT	2,898	0.053	-0.406	0.169
WBC	All	*CXCR2*	VT	2	5	138,814	3.56E-13	-0.21	0.028	0.0023	10,438					
WBC	All	*JAK2*	VT	9	6	138,814	2.682E-06	0.334	0.066	0.00036	8,045	VT	2,898	0.0091	-0.505	0.163
WBC	**EA**	***TAF3***	**VT**	**10**	**6**	**108,596**	**1.583E-06**	**0.328**	**0.064**	**0.00037**	**6,848**	**VT**	2,898	**NA**	**NA**	**NA**
NEU	EA	*CXCR2*	SKAT	2	8	60,851	6.853E-09	-0.221	0.038	1	2,950	SKAT	1,285	0.190	-0.017	0.236
NEU	All	*CXCR2*	SKAT	2	7	77,223	5.062E-09	-0.203	0.035	1	9,252					
NEU	EA	*CXCR2*	VT	2	8	60,851	2.495E-08	-0.221	0.038	0.0185	2,950	VT	1,285	0.580	-0.392	0.335
NEU	All	*CXCR2*	VT	2	4	77,223	2.128E-08	-0.218	0.037	0.0023	9,252					
NEU	All	*JAK2*	VT	9	6	77,223	1.553E-07	0.431	0.077	0.00054	6,388	VT	1,285	NA	NA	NA
NEU	**EA**	***OR4C6***	**SKAT**	**11**	**15**	**60,851**	**2.561E-08**	**0.109**	**0.044**	**1**	**512**	**SKAT**	**1,285**	**0.4844**	**-0.1455**	**0.198**
NEU	EA	*IL17RA*	SKAT	22	18	44,325	1.004E-16	-0.128	0.026	1	10,016	SKAT	1,285	0.949	-0.018	0.176
NEU	**AA**	***ZNF439***	**VT**	**19**	**4**	**10,119**	**9.566E-07**	**-1.099**	**0.221**	**0.083**	**1,688**	**VT**	1,285	**NA**	**NA**	**NA**
MON	All	*IL17RA*	SKAT	22	19	57,567	1.174E-18	-0.104	0.023	1	23,595	SKAT	1,285	0.972	-0.004	0.176
LYM	**All**	***TBX3***	**VT**	**12**	**4**	**60,363**	**1.959E-06**	**0.419**	**0.083**	**0.00060**	**144**	**VT**	**1,285**	**NA**	**NA**	**NA**

Potentially novel genes are shown in bold.

Abbreviations: Chr, Chromosome; MAF, minor allele frequency; cMAC, cumulative minor allele count; WBC, white blood cell; NEU, neutrophil; MON, monocytes; LYM, lymphocytes; SKAT, sequence kernel association test; VT, variable-threshold test; ; EA, European ancestry; AA, African ancestry; All, combined European, African, Hispanic American, East Asian and South Asian ancestries.

Table S11. Association between white blood cell trait associated variants and immunologically relevant quantitative traits in previous genome-wide association studies (P < 1.64E-04).

Trait (population)	dbSNPID	Chr	Pos	Alt/Ref	Gene	Quantitative traits	Sample size	P[a]	PMID
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	C-reactive protein	70,410	3.80E-43	23263486, 21300955, 23505291, 22939635, 18439548
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	Plasma protein-C levels	9,424	2.00E-17	20802025
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	Serum urate	50,337	5.90E-17	20884846, 23263486, 21768215, 19503597, 21943158
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	Platelet count	37,438	9.10E-10	22139419
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	2-hour glucose tolerance test	133,010	9.00E-15	22885924, 20081857, 23263486
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	Fasting blood glucose	133,010	2.20E-41	22885924, 20081857, 20081858, 23263486
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	Fasting insulin	133,010	2.70E-22	22885924, 20081857, 20081858, 23263486
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	Factor VII activity in plasma	31,212	6.20E-24	21676895, 20231535, 21676895
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	Total cholesterol	140,059	4.40E-28	20686565, 23063622, 20339536, 20339536, 19060906
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	Triglycerides	140,059	1.30E-139	20686565, 22629316, 23063622, 19936222, 18193043, 20657596, 19060911, 19802338, 20339536, 23505323, 20139978, 19060910, 23236364, 21943158, 18454146, 19913121, 21862451, 19060906, 22171074
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	HDL	26,768	6.30E-36	19936222
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	Height	183,727	9.40E-05	20881960
WBC (EA)	rs1260326	2	27,730,940	C/T	*GCKR*	HOMA-IR	122,744	9.20E-07	20081858, 20081857, 23263486
WBC (All)	rs185819	6	32,050,067	C/T	*TNXB*	Height	183,727	1.70E-06	20881960, 18391951
WBC (All)	rs9374080	6	109,616,420	C/T	*CCDC162P*	Mean corpuscular hemoglobin concentration	135,367	3.00E-21	23222517
WBC (All)	rs9374080	6	109,616,420	C/T	*CCDC162P*	Mean corpuscular volume	135,367	2.30E-18	23222517, 19862010, 20139978
WBC (All)	rs9374080	6	109,616,420	C/T	*CCDC162P*	Red blood cell count	135,367	1.60E-15	23222517
MON (All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	HDL	140,059	2.40E-05	20686565
WBC (EA)	rs1935	10	64,927,823	G/C	*JMJD1C*	Platelet count	16,388	2.70E-08	22423221
WBC (EA)	rs1935	10	64,927,823	G/C	*JMJD1C*	Triglycerides	140,059	1.50E-07	20686565, 19060906
MON (EA)	rs11625112[b]	14	23,596,740	G/A	*SLC7A8*	Triglycerides	140,059	4.90E-05	20686565, 19060906
WBC (All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	C-reactive protein	82,725	4.00E-08	21300955, 22939635

[a]Lookup of 16 single nucleotide variants was performed and results with multiple testing corrected p-value < 1.64E-04 (16 variants and 19 quantitative traits) are shown here. When multiple studies report the same variant-trait associations, results from the largest sample size are presented here.

[b] LD r^2 between rs11625112 and rs2239633 in *SLC7A8* is 0.63.

Abbreviations: Chr, chromosome; Pos, basepair position; Alt, effect allele; Ref, reference allele; HDL, high density lipoprotein; HOMA-IR, homeostatic model assessment-insulin resistance; WBC, white blood cell; NEU, neutrophil; MON, monocyte; LYM, lymphocyte; BAS, basophil.

Table S12. White blood cell trait variants associated with platelet and red blood cell related traits in the European and combined All ancestries of the Blood-Cell Consortium.

dbSNPID	Chr	Pos	Gene	Alt/Ref	EAF	Trait	Beta	SE	P
European ancestry									
rs1260326	2	27,730,940	*GCKR*	C/T	0.58	WBC	-0.03	0.005	4.01E-10
						PLT	-0.037	0.005	7.15E-14
						RBC	0.022	0.006	5.73E-04
						HCT	0.023	0.005	7.53E-06
						HGB	0.022	0.005	3.74E-06
rs2229094	6	31,540,556	*LTA*	C/T	0.26	LYM	0.046	0.008	1.89E-08
						PLT	0.019	0.006	2.04E-03
rs1935	10	64,927,823	*JMJD1C*	G/C	0.49	WBC	-0.026	0.005	2.46E-08
						PLT	0.05	0.005	3.11E-25
						MPV	-0.112	0.008	7.64E-41
						RDW	-0.017	0.007	1.75E-02
						MCH	0.017	0.007	1.18E-02
rs4763879	12	9,910,164	*CD69*	A/G	0.36	LYM	-0.038	0.007	3.08E-08
						HGB	-0.01	0.005	3.44E-02
						MCHC	-0.012	0.006	3.38E-02
rs11625112	14	23,596,740	*SLC7A8*	G/A	0.46	MON	-0.038	0.007	3.82E-08
						HCT	0.012	0.005	1.76E-02
rs1292053	17	57,963,537	*TUBD1*	G/A	0.45	WBC	-0.03	0.004	1.28E-11
						PLT	-0.017	0.005	1.94E-04
						RDW	0.024	0.007	3.85E-04
						HCT	-0.015	0.005	2.05E-03
						HGB	-0.016	0.005	5.11E-04
						MCV	-0.012	0.006	4.34E-02
						MCH	-0.016	0.006	1.45E-02
rs3865444	19	51,727,962	*CD33*	A/C	0.31	WBC	-0.037	0.005	3.51E-12
						PLT	-0.026	0.005	1.11E-06
						MCHC	0.016	0.006	8.70E-03
rs2295764	20	31,025,163	*ASXL1*	G/A	0.36	BAS	-0.042	0.007	3.28E-09
						RDW	0.034	0.007	2.35E-06
All combined ancestries									
rs2276853	3	47,282,303	*KIF9*	A/G	0.58	WBC	0.023	0.004	3.65E-08
						HGB	-0.011	0.004	1.14E-02
rs9374080	6	109,616,420	*CCDC162P*	C/T	0.43	WBC	0.023	0.004	4.01E-08
						PLT	0.015	0.004	3.07E-04
						MPV	-0.024	0.007	9.28E-04
						RBC	-0.043	0.005	3.13E-15
						RDW	-0.041	0.006	6.19E-11

						MCV	0.06	0.006	8.84E-28
						MCH	0.062	0.006	3.10E-26
						MCHC	0.023	0.005	5.13E-06
rs185819	6	32,050,067	*TNXB*	C/T	0.51	WBC	0.031	0.005	4.02E-10
						PLT	0.017	0.005	6.16E-04
						MPV	-0.023	0.009	8.78E-03
						MCH	-0.015	0.007	2.94E-02
rs1292053	17	57,963,537	*TUBD1*	G/A	0.45	MON	-0.036	0.006	2.55E-09
						PLT	-0.013	0.004	1.40E-03
						RBC	-0.012	0.005	2.51E-02
						RDW	0.023	0.006	2.30E-04
						HCT	-0.013	0.004	2.97E-03
						HGB	-0.013	0.004	1.02E-03
rs2836878	21	40,465,534	*ETS2-PSMG1*	A/G	0.26	WBC	-0.025	0.005	8.36E-08
						PLT	-0.014	0.005	3.44E-03
						RDW	-0.025	0.007	3.28E-04
						HCT	0.016	0.005	1.37E-03
						HGB	0.022	0.005	2.72E-06
						MCV	0.014	0.006	2.14E-02
						MCH	0.024	0.007	2.78E-04
						MCHC	0.019	0.006	9.14E-04

Abbreviations: Chr, chromosome; Pos, basepair position; Alt, effect allele; Ref, reference allele; EAF, effect allele frequency; EA, European ancestry, All, combined European, African, Hispanic American, East Asian and South Asian ancestries; BAS, basophil; HCT, hematocrit; HGB, hemoglobin; LYM, lymphocyte; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; MON, monocytes; MPV, mean platelet volume; PLT, platelet; RBC, red blood cell; RDW, red cell distribution width; WBC, white blood cell.

See Table S13 in the accompanying Excel file.

Additional Funding Information

Airwave

The Airwave Study is funded by the Home Office (grant number 780-TETRA) with additional support from the National Institute for Health Research (NIHR) Imperial College Healthcare NHS Trust (ICHNT) and Imperial College Biomedical Research Centre (BRC) (Grant number BRC-P38084). Paul Elliott is an NIHR Senior Investigator and is supported by the ICHNT and Imperial College BRC, the MRC-PHE Centre for Environment and Health and the NIHR Health Protection Research Unit on Health Impact of Environmental Hazards.

ARIC

The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C.

BioMe

The Mount Sinai Institute for Personalized Medicine Biobank Program is supported by The Andrea and Charles Bronfman Philanthropies.

BIOVU

The dataset used in the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS (National Center for Advancing Translational Sciences)/National Institutes of Health (NIH). Genome-wide genotyping was funded by NIH grants RC2GM092618 from NIGMS/OD and U01HG004603 from the National

Human Genome Research Institute (NHGRI)/ National Institute of General Medical Sciences (NIGMS). Funding for TLE and DRVE was provided by 1R21HL12142902 from NHLBI/NIH. Funding for the BioVU replication cohort was provided by 5R01HD074711 from National Institute of Child Health and Human Development (NICHD)/NIH. PheWAS studies were supported by R01LM010685 from the National Library of Medicine.

CARDIA

The CARDIA Study is conducted and supported by the NHLBI in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging (NIA). Exomechip genotyping was supported from grants R01-HL093029 and U01- HG004729 to MF. This manuscript has been reviewed and approved by CARDIA for scientific content.

CHS

This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants HL080295, HL087652, HL103612, HL105756, HL120393, R01HL068986 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629 from the NIA. A full list of CHS investigators and institutions can be found at http://chs-nhlbi.org/. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

EGCUT

This study was supported by EU H2020 grants 692145, 676550, 654248, Estonian Research Council Grant IUT20-60, NIASC, EIT – Health and NIH-BMI grant 2R01DK075787-06A1.

FINCAVAS

This work was supported by the Competitive Research Funding of the Tampere University Hospital (Grant 9M048 and 9N035), the Finnish Cultural Foundation, the Finnish Foundation for Cardiovascular Research, the Emil Aaltonen Foundation, Finland, and the Tampere Tuberculosis Foundation.

FramHS

Genotyping, quality control and calling of the Illumina HumanExome BeadChip in the Framingham Heart Study was supported by funding from the National Heart, Lung and Blood Institute Division of Intramural Research (Daniel Levy and Christopher J. O'Donnell, Principle Investigators). Support for the centralized genotype calling was provided by Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium through the NIH American Recovery and Reinvestment Act of 2009 (5RC2HL102419). The NHLBI's Framingham Heart Study is a joint project of the National Institutes of Health and Boston University School of Medicine and was supported by contract N01-HC-25195.

GeneSTAR

GeneSTAR was supported by the NIH/NHLBI (U01 HL72518, HL087698, and HL112064) and by a grant from the NIH/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center. Genotyping services were provided through the RS&G Service by the Northwest Genomics Center at the University of Washington, Department of Genome Sciences, under U.S. Federal Government contract number HHSN268201100037C from the National Heart, Lung, and Blood Institute.

HABC

HABC funding/acknowledgement: The Health ABC Study was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106 and, in part, by the NIA Intramural Research Program. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov).

HANDLS

The Healthy Aging in Neighborhoods of Diversity across the Life Span Study (HANDLS) research was supported by the Intramural Research Program of the NIH, NIA and the National Center on Minority Health and Health Disparities (project # Z01-AG000513 and human subjects protocol # 2009-149). Data analyses for the HANDLS study utilized the computational resources of the NIH HPC Biowulf cluster at the National Institutes of Health, Bethesda, MD (http://hpc.nih.gov).

JHS

We thank the Jackson Heart Study (JHS) participants and staff for their contributions to this work. The JHS is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparities.

LBC1921 and LBC1936

Phenotype collection in the Lothian Birth Cohort 1921 was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Genotyping was supported by Centre for Cognitive Ageing and Cognitive

Epidemiology (Pilot Fund award), Age UK, and the Royal Society of Edinburgh. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged. WDH is supported by a grant from Age UK (Disconnected Mind Project).

MESA

MESA and the MESA SHARe project are conducted and supported by the NHLBI in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-001079, UL1-TR-000040, and DK063491. MESA Family is conducted and supported by the NHLBI in collaboration with MESA investigators. Support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, by the National Center for Research Resources, Grant UL1RR033176, and the NCATS, Grant UL1TR000124. Funding support for the inflammation dataset was provided by grant HL077449. The MESA Epigenomics & Transcriptomics Study was funded by NIA grant 1R01HL101250-01 to Wake Forest University Health Sciences.

MHIBB

This work was supported by the Fonds de Recherche du Québec–Santé (FRQS, scholarship to NC), the Canadian Institute of Health Research (Banting-CIHR, scholarship to SL and operating grant MOP#123382 to GL), the Canada Research Chair program (to GL, JDR, and JCT), and the Montreal Heart Institute (MHI) Foundation. JCT holds the Canada Research Chair in translational and personalized medicine and the Université de Montréal endowed research chair in atherosclerosis.

REGARDS

The genotyping for this project was provided by NIH/National Center for Research Resources (NCRR) grant 5U54RR026137-03.

RS

The generation and management of the Illumina Exomechip v1.0 array data for the Rotterdam Study (RS-I) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The Exomechip array data set was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, from the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO)-sponsored Netherlands Consortium for Healthy Aging (NCHA; project nr. 050-060-810); the Netherlands Organization for Scientific Research (NWO; project number 184021007) and by the Rainbow Project (RP10; Netherlands Exomechip Project) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL; www.bbmri.nl). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

SHIP and SHIP-Trend

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network ‘Greifswald Approach to Individualized Medicine (GANI_MED)’ funded by the Federal Ministry of Education and Research (grant 03IS2061A). ExomeChip data have been supported by the Federal Ministry of Education and Research (grant no. 03Z1CN22) and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the ‘Center of Knowledge Interchange’ program of the Siemens AG and the Caché Campus program of the InterSystems GmbH. The SHIP and SHIP-TREND samples were genotyped at the Helmholtz Zentrum München.

STABILITY and SOLID TIMI-52

The STABILITY and SOLID TIMI-52 studies were funded by GlaxoSmithKline. Michelle L. O'Donoghue reports receiving grants from GlaxoSmithKline, Merck, AstraZeneca, and Eisa.

WHI

The WHI program is funded by the NHLBI, NIH, and the US Department of Health and Human Services (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). Exomechip data and analysis were supported through the Exome Sequencing Project (NHLBI RC2 HL-102924, RC2 HL-102925 and RC2 HL-102926), the Genetics and Epidemiology of Colorectal Cancer Consortium (National Cancer Institute CA137088), and the Genomics and Randomized Trials Network (NHGRI U01-HG005152). A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf"

Supplemental Acknowledgments

The MHI Biobank thanks all participants and staff of the MHI Biobank and acknowledges the technical support of the Beaulieu-Saucier MHI Pharmacogenomic Center. The WHI study thanks investigators and staff for their dedication in making the program possible. The FHS authors are pleased to acknowledge the Shared Computing Cluster, which is administered by Boston University's Research Computing Services. URL: www.bu.edu/tech/support/research/. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services. ARIC thanks the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The meta-analysis and meta-regression analyses were funded by grant R01 HL086694 from the National Heart, Lung, and Blood Institute. Airwave study thanks all participants in the Airwave Health Monitoring Study;

Louisa Cavaliero who assisted in data collection and management as well as Peter McFarlane and the Glasgow CARE, Patricia Munroe at Queen Mary University of London, Joanna Sarnecka and Ania Zawodniak at Northwick Park for their contributions to the study. GeneSTAR thanks all of our participating families. MESA thanks its Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org. A full list of CHS investigators and institutions can be found at http://chs-nhlbi.org/. SOLID-TIMI-52 and STABILITY thank Liling Warren for contributions to the genetic analysis of the study datasets. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the InterSystems GmbH. The SHIP and SHIP-TREND samples were genotyped at the Helmholtz Zentrum München, Germany. EGCUT thanks all the participants and co-workers at Estonian Biobank, especially Mr. V. Soo, Mr. S. Smith and Dr. L. Milani. FINCAVAS thanks the staff at the Department of Clinical Physiology for data collection. LBC thanks the cohort participants and team members who contributed to these studies. RS thanks Ms. Mila Jhamai, Ms. Sarah Higgins, and Mr. Marijn Verkerk for their help in creating the exomechip database, and Carolina Medina-Gomez, MSc, Lennard Karsten, MSc, and Linda Broer Ph.D. for quality control and variant. RS is also grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. REGARDS thanks the investigators, staff and participants of the study for their valuable contributions. A full list of participating REGARDS investigators and institutions can be found at http://www.regardsstudy.org.