Large-scale exome-wide association analysis identifies loci for White Blood Cell Traits and Pleiotropy with Immune-Mediated Diseases

Citation for published version:
Tajuddin, SM, Schick, UM, Eicher, JD, Chami, N, Giri, A, Brody, JA, Hill, WD, Kacprowski, T, Li, J, Lyytikäinen, L-P, Manichaikul, A, Mihailov, E, O'Donoghue, ML, Pankratz, N, Pazoki, R, Polfus, LM, Smith, AV, Schurmann, C, Vacchi-Suzzi, C, Waterworth, DM, Evangelou, E, Yanek, LR, Burt, A, Chen, M-H, van Rooij, FJA, Floyd, JS, Greinacher, A, Harris, TB, Highland, HM, Lange, LA, Liu, Y, Mägi, R, Nalls, MA, Mathias, RA, Nickerson, DA, Nikus, K, Starr, JM, Tardif, J-C, Tzoulaki, I, Velez Edwards, DR, Wallentin, L, Bartz, TM, Becker, LC, Denny, JC, Raffield, LM, Rioux, JD, Friedrich, N, Fornage, M, Gao, H, Hirschhorn, JN, Liewald, DCM, Rich, SS, Uitterlinden, A, Bastarache, L, Becker, DM, Boerwinkle, E, de Denus, S, Bottinger, EP, Hayward, C, Hofman, A, Homuth, G, Lange, E, Launer, LJ, Lehtimäki, T, Lu, Y, Metspalu, A, O'Donnell, CJ, Quarells, RC, Richard, M, Torstenson, ES, Taylor, KD, Vergnaud, A-C, Zonderman, AB, Crosslin, DR, Deary, IJ, Dörr, M, Elliott, P, Evans, MK, Gudnason, V, Kähönen, M, Psaty, BM, Rotter, JI, Slater, AJ, Dehghan, A, White, HD, Ganesh, SK, Loos, RJF, Esko, T, Faraday, N, Wilson, JG, Cushman, M, Johnson, AD, Edwards, TL, Zakai, NA, Lettre, G, Reiner, AP & Auer, PL 2016, 'Large-scale exome-wide association analysis identifies loci for White Blood Cell Traits and Pleiotropy with Immune-Mediated Diseases', *American Journal of Human Genetics*, vol. 99, no. 1, pp. 22-39. https://doi.org/10.1016/j.ajhg.2016.05.003

Digital Object Identifier (DOI):
10.1016/j.ajhg.2016.05.003

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
American Journal of Human Genetics

Large-scale exome-wide association analysis identifies loci for white blood cell traits and pleiotropy with immune-mediated diseases

Salman M. Tajuddin,[1,*], Ursula M. Schick,[2,3,*] 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 Polfus,[21] Albert Vernon Smith,[22,23] Claudia Schurmann,[2] 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] Jean-Claude Tardif,[5,6] Ioanna Tzoulaki,[26] Digna R. Velez Edwards,[40] Lars Wallentin,[41] Traci M. Bartz,[42] Lewis C. Becker,[43] Joshua C. Denny,[44] Laura M. Raffield,[33] John D. Rioux,[5,6] Nele Friedrich,[12,45] Myriam Fornage,[46] He Gao,[26] Joel N. Hirschhorn,[47,48] David C. M. Liewald,[9,10] Stephen S. Rich,[16] Andre Uitterlinden,[20] Lisa Bartaracha,[44] Diane M. Becker,[28] Eric Boerwinkle,[21,49] Simon de Denus,[6,50] Erwin P. Bottinger,[2], Caroline Hayward,[51] Jussi Hernesniemi,[14,15] Albert Hofman,[20,52] Georg Homuth,[11] Ethan Lange,[53] Lenore J. Launer,[31] Yingchang Lu,[3,54] Andres Metspalu,[17] Chris J. O'Donnell,[55,56] Rakale C. Quarells,[57] Melissa Richard,[46] Eric S. Torstenson,[7] Kent D. Taylor,[58,59] Anne-Claire Vergnaud,[26] Alan B. Zonderman,[1] David R. Crosslin,[60] Ian J. Deary,[9,10] Marcus Dörr,[12,61] Paul Elliott,[26] Michele K. Evans,[1] Vilmundur Gudnason,[22,23] Mika Kähönen,[62,63] Bruce M. Psaty,[64] Jerome I. Rotter,[58,59] Andrew J. Slater,[65] Abbas Dehghan,[20] Harvey D. White,[66] Santhi K. Ganesh,[67] Ruth J. F. Loos,[2] Tõnu Esko,[17,47] Nauder Faraday,[68] James G. Wilson,[69] Mary Cushman,[70] Andrew D. Johnson,[4] Todd L. Edwards,[71] Neil A. Zakai,[70] Guillaume Lettre,[5,6,*] Alex P. Reiner,[72,73,*] Paul L. Auer,[74,*]

Affiliations

[1] Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, 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, Québec, H3T 1J4, Canada

[6] Montreal Heart Institute, Montréal, Québec, 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 Institue for Genetics and Functional Genomics, University Medicine Greifswald and Ernst-Mortiz-Arndt University, Greifswald, 17489, Germany

[12] DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald, 13347, 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, 33521, Finland

[15] Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, 33520, 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] 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 MC, Rotterdam, 3000, 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, IS-201 Kopavogur, Iceland

[23] Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland

[24] Department of Family, 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, National Institutes of Health, 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, National Institutes of Health, 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 Hospital, Tampere University Hospital, Tampere, 33521, Finland

[39] Department of Cardiology, University of Tampere School of Medicine, Tampere, 33520, Finland

[40] Vanderbilt Epidemiology Center, Department of Obstetrics & Gynecology, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN 37235, USA

[41] Department of Medical Sciences, Cardiology, and Uppsala Clinical Research Center, Uppsala University, SE-751 83, Uppsala, Sweden

[42] Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

[43] Department of Medicine, Divisions of Cardiology and General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

[44] Department of Biomedical Informatics, School of Medicine, Vanderbilt University, Nashville, TN 37203, USA

[45] Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, 13347, Germany

[46] Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

[47] Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA

[48] Department of Endocrinology, Boston Children's Hospital, Boston, MA 02115, USA

[49] Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

[50] Faculty of Pharmacy, Université de Montréal, Montréal, Québec, H3T 1J4, Canada

[51] Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK

[52] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[53] Departments of Genetics and Biostatistics, University of North Carolina at Chapel Hill, University of North Carolina, Chapel Hill, NC 27599, USA

[54] The Charles Bronfman Institute for Personalized Medicine, The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[55] National Heart Lung and Blood Institute, The Framingham Heart Study, Framingham, MA 01702, USA

[56] Cardiology Section and Center for Population Genomics, Boston Veteran's Administration (VA) Healthcare, Boston, MA 02118, USA

[57] Morehouse School of Medicine, Social Epidemiology Research Center, Cardiovascular Research Institute, Atlanta, GA 30310, USA

[58] Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute, Torrance, CA 90502, USA

[59] Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA 90502, USA

[60] Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA 98195, USA

[61] Department of Cardiology, University Medicine Greifswald, Greifswald, 17475, Germany

[62] Department of Clinical Physiology, Tampere University Hospital, Tampere, 33521, Finland

[63] Department of Clinical Physiology, University of Tampere School of Medicine, Tampere, 33520, Finland

[64] Departments of Epidemiology, Health Services, and Medicine, University of Washington, Seattle, WA 98195, USA

[65] OmicSoft Corporation, Cary, NC 27513, USA

[66] Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, 1142, New Zealand

[67] Departments of Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, MI 48108, USA

[68] Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

[69] Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA

[70] Division of Hematology Oncology, Department of Medicine, The University of Vermont, Colchester, VT 05446, USA

[71] Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN 37203, USA

[72] Department of Epidemiology, University of Washington, Seattle, WA 98195, USA

[73] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

[74] Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI 53205, USA

*These authors contributed equally to this study.

Correspondence: A.P.R. (apreiner@u.washington.edu) and P.L.A. (pauer@uwm.edu)

Abstract

White blood cells play diverse roles in innate and adaptive immunity. Genetic association analyses of phenotypic variation in circulating white blood cell counts (WBC) from large samples of otherwise healthy individuals can provide insights into genes or 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 (eight 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*), 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 (WBC) 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 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 (GWAS) have characterized common and lower frequency variation contributing to WBC counts in European-, African- and Asian-ancestry populations.[3-13] More than 30 distinct genetic loci have been discovered; in some instances, these genetic studies have provided important new 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 WBC traits in European-ancestry populations[6] and less than 25% in African-ancestry (AAs) populations (in AAs, a substantial proportion of the variation in WBC counts is attributed to a single variant rs2814778 in the *DARC* (Duffy Antigen Receptor for Chemokines, MIM 613665) gene).[3, 14] In an effort to augment the discoveries from GWAS and to identify additional functional loci contributing to variation in WBC counts, we performed the largest 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 (**Table 1** and **Table S2-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 only participate 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 chip genotyping arrays: Illumina ExomeChip v1.0, Illumina ExomeChip v1.1_A, Illumina ExomeChip-12 v1.1, Affymetrix Axiom Biobank Plus GSKBB1, Illumina HumanOmniExpressExome Chip. Genotypes were called either 1) using a combination of the Illumina GenomeStudio and zCall software or 2) the exome chip joint calling plan developed by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium[15] (**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. Following all quality control procedures, each study generated an indexed variant call file (VCF) for subsequent analyses. The VCF files were checked for allele alignment using the checkVCF package.

We performed study specific quality control on each trait association results using the EasyQC protocol.[16] 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 scatter plot 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 exome chip 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.[17] Autosomal and X chromosome variants were then tested for association with each WBC traits 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 1MB 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.[18] We included variants in the meta-analysis if the genotype call rate was ≥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.[16] 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 have opposing direction of effects.[18] For all gene-based tests performed, we considered single nucleotide variants (SNVs) with an allele frequency of ≤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 association signals within a region, using the RareMETALS package we performed step-wise conditional analyses adjusting for the most significant single variant in a 1 MB window, across the entire exome-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 step-wise conditional analysis in GCTA software[19] 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 4 independent European ancestry cohorts (**Table 1, Table 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.[20] We also performed replication of gene-based associations in independent ~2,900 EA samples.

Phenome-wide Association Study (PheWAS) Analysis

In 29,722 EA samples from the BioVU study,[21] we performed PheWAS analysis[22] to assess the association between our WBC related loci and 1,502 International Classification of Disease, Ninth Revision (ICD-9) code curated clinical phenotypes.[22] Variants were included in the analysis if there were 10 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 GWAS studies 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).[23-31] We supplemented the full GWAS summary statistics lookups with the GRASP database[32] 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.[33, 34]

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 over 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,[35] with specific citations for the blood cell specific eQTL datasets shown in **Table S5**. Additional in silico functional annotations were performed using ANNOVAR.[36] The deleteriousness of each variant was estimated using 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.[37]

Results

We conducted an exome-wide association analyses of total WBC and differential counts (neutrophils, monocytes, 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 (**Table 1**, **Table 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 ($p\text{-value} < 2.0 \times 10^{-7}$) (**Table S6**). Following step-wise conditional analyses, we refined this list to 28 independent SNV associations with WBC count that were not previously reported (**Table S7**). Of these 28 variant associations, 16 were replicated ($p\text{-value} < 0.05$ 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 1MB window of a previously reported WBC locus. Of the 16 replicated loci, ten were significantly associated with total WBC count, two with neutrophil count, four with monocyte count, two with lymphocyte count, and one 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 exome chip 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) was associated with lower total WBC in the EA meta-analysis (p-value = 8.13×10^{-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) was associated with increased total WBC in the multi-ancestry meta-analysis (p-value = 3.29×10^{-9}). The signal was largely driven 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) was associated with increased total WBC 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) 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) 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) was associated with lower total WBC in the

EA meta-analysis (p-value = 6.55 x 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) variant was associated with lower total WBC (p-value = 8.34 x 10^{-13}) in the EA meta-analysis. The effect allele frequencies were highly discrepant across ancestries, perhaps explaining why the association was only observed 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 x 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 from *CD33* (MIM 159590), was associated with lower total WBC in the EA meta-analysis (p-value = 6.81 x 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 x 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, EAF = 0.00042) that was significantly associated with lower total WBC in the EA discovery analysis (p-value = 1.87 x 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 x 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 x 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 x 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 x 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 x 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 x 10^{-9}). We also found a secondary signal rs11642873 near *IRF8* (MIM 601565)[38] that was associated with higher monocyte count in the EA meta-analysis [discovery beta (p-value) = 0.072 (1.40 x 10^{-22}), conditional beta (p-value) = 0.054 (1.41 x 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 x 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) in *LTA* (MIM 153440) was associated with higher lymphocyte count in the EA meta-analysis (p-value = 3.14 x 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, EAF in EAs = 5.25 x $10^{-5,}$ discovery p-value = 7.56 x 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 *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}) (**Table 3, Table S11**).

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. While 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.[39] 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 exome chip effort, 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 10 additional exome chip identified loci we report is 0.0183. Thus, our exome chip 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 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.[40] *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 which results in lower full-length *CD33* expression due to skipping of exon 2.[41]

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.[42, 43] The deleterious coding variant of *CD87* rs4760A>G (p.Leu272Pro)[37] 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 and migration, chemotaxis, and a regulator of the uptake by macrophages of apoptotic neutrophils.[44] It is possible that the latter mechanism might explain the selective association of rs4760 with lower neutrophil count.

The *CD69* intronic allele rs4763879 G>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.[45] 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).[45] 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,[46] - and with platelet traits (**Table S11-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.[47] *CD164* regulates CXCR4/CXCL12 signaling in hematopoietic precursor cells.[48] 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.[49] 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 (**Table S12, S13**). *ASXL1* is a chromatin binding transcriptional regulator of the polycomb group and hematopoietic tumor suppressor gene.[50] *JMJD1C* is also an epigenetic regulator of gene expression, likely through histone demethylation.[51] The association between *JMJD1C* and lower WBC counts (this study), platelet count, mean platelet volume, and platelet reactivity[52] indicate multi-lineage effects on hematopoiesis. *JMJD1C* was originally identified as a ligand-dependent interacting partner of thyroid hormone and androgen receptors.[53] In human myeloid leukemia cells, *JMJD1C* functions as a coactivator for the leukemogenic transcriptional complex *RUNX1-RUNX1T1* to increase AML cell proliferation and survival.[54] 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, while 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.[55]

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 granulpoiesis.[56] *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*.[57]

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.[58] 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.[59] Recent data also suggest a role of *CEBPE* isoforms in differential regulation of eosinophil production as well as in the monocyte-granulocyte lineage decision.[60]

The transcription factor encoded by *IKZF1* or Ikaros was initially described as a regulator of lymphoid lineage differentiation and hematopoietic progenitor cell self-renewal.[61] An Ikaros isoform selectively expressed in myeloid precursor cells was subsequently found to regulate myeloid differentiation.[61] 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*.[49] 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 involving in regulation of cell death and apoptosis

Apoptosis regulates hematopoietic stem cells, and maintains the balance between cell proliferation and cell death.[62] Altered apoptotic processes contribute to the development of autoimmune and other inflammatory diseases.[63] 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, *DD1α*), 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.[64] *DD1α* 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.[65]

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.[66] Loss of *LTA* was associated with a four-fold increase in B lymphocytes in peripheral blood count in mice.[66] The index missense SNP is also a cis-eQTL for *LTA* and *NFKBIL1* (MIM 601022) (**Table S13**).

Loci involving 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 may 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.[67] The index SNP localizes to the major histocompatibility complex class III region on chromosome 6 and overlaps the *ATF6B* and *CYP21A2* genes 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[68], is also a key virulence factor for *Toxoplasma gondii*.[69]

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.[70] *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* (**Table S12, 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.[71] The PI3K pathway is also involved in Toll-like receptor (TLR) signaling and release of cytokines from macrophages,[72] and a proxy SNP of

TUBD1 rs1292053 has been associated with CRP.[73] *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*.[49]

The *GCKR* rs1260326 variant is an eQTL for (1) *SNX17*, which has been associated with T cell activation and is a binding protein for human papillomavirus L2 capsid protein; and (2) *NRBP1*, which binds a Dengue virus protein.[49, 74]

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.[75] Our results add to recent evidence that genetic factors controlling WBC and immune cell counts contribute to autoimmune disease risk.[76] Several loci involve regulation of cellular mechanisms critical in the development of autoimmune diseases such as modulation of autoimmune reactivity (*CD69*),[77] and apoptosis (*LTA*, *DD1a*, *CD87*).[44, 64]

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's disease, and Stevens-Johnson syndrome (**Figure 2, Table 3**). While 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, epithelialassociated stimulated dendritic cell genes in Crohn disease; and monocyte-specific eQTLs among neurodegenerative disease variants.[78, 79]

Abnormal inflammatory response and activation of microglial cells are linked with the development of AD and other neurodegenerative diseases. The WBC-associated *CD33* gene is among the inflammation related AD risk loci identified by GWAS studies.[80] A variant in this gene was shown to modulate *CD33* exon 2 splicing efficiency leading to abnormal activation of microglial cells which are tissue-resident macrophages of the brain derived from monocyte lineage cells.[80] 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*) which functions as growth factor for hematopoietic progenitor cells.[81] Several of the same loci are involved in susceptibility to infectious diseases (*IRF8* and mendelian susceptibility to mycobacterial disease (MIM 209950),[82] *TNXB* associated with *T. gonadii* and climatic adaptation,[69, 83] malaria with *ABO* (MIM 110300) and *DARC*,[3, 84] *CD87* and clearance of bacteria[85]), 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.[86] 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 target for anti-leukemic therapy.[41] The *CD33* rs3865444 and rs12459419 variants associated with lower WBC count and alternative splicing of exon 2, respectively have been associated with both Alzheimer's disease risk and AML treatment efficacy.[41] 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.[41, 87] *CD87* is expressed on various immune cells including, neutrophils, monocytes, macrophages, T cells, and basophils, as well as endothelial cells and hepatocytes.[42, 88] The cleaved soluble form of CD87 may have a role in hematopoietic stem/progenitor cell mobilization.[89]

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).[50, 90, 91] Knockdown of *Asxl1* in mouse results in impaired lymphoid and myeloid differentiation and multi-lineage cytopenias.[92] 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.[55] Recent studies have shown that *IKZF1* deletions and mutations that caused reduction of Ikaros activity are highly associated with development of acute lymphoblastic leukemia.[93, 94] 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.[53] 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*).[13] 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.[95] *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[96] whereas *GATA3* variants have been associated with susceptibility to hematologic malignancies.[97] 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, likely 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 7 different phenotypes in 3 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.[39] 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 two figures and 13 tables.

Acknowledgments

We thank all participants and study coordinating centers of the participating studies and cohorts. PLA was supported by NHLBI R21 HL121422-02. GL 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, National Institutes of Health Intramural Research Program. Data analyses utilized the computational resources of the NIH HPC Biowulf cluster at the National Institutes of Health, Bethesda, MD (http://hpc.nih.gov). 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, Joanna Sarnecka and Ania Zawodniak at Northwick Park for their contributions to the study. SOLID-TIMI-52 and GSK-STABILITY thank Liling Warren for contributions to the genetic analysis of the study datasets. EGCUT thanks Mr. V. Soo, Mr. S. Smith and Dr. L. Milani for their contribution. RS thanks Ms. Mila Jhamai, Ms. Sarah Higgins, and Mr. Marijn Verkerk for their help in creating the exome chip database, and Ms. Carolina Medina-Gomez, Mr. Lennard Karsten, and Dr. Linda Broer. Additional acknowledgments and funding information are provided in the **Supplemental Data**.

Web Resources

CheckVCF, https://github.com/zhanxw/checkVCF

GRASP, http://grasp.nhlbi.nih.gov/Overview.aspx

HPC Biowulf Linux cluster, https://hpc.nih.gov/

Online Mendelian Inheritance in Man (OMIM), http://www.omim.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

The 1000 Genomes Project, http://www.1000genomes.org/

The R Project for Statistical Computing, https://www.r-project.org/

White blood cell traits full exome chip summary association results, http://www.mhi-humangenetics.org/en/resources

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. (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. (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. (2014). Trans-ethnic meta-analysis of white blood cell phenotypes. Hum Mol Genet 23, 6944-6960.

9. Soranzo, N., Spector, T.D., Mangino, M., Kuhnel, B., Rendon, A., Teumer, A., Willenborg, C., Wright, B., Chen, L., Li, M., et al. (2009). A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet 41, 1182-1190.

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., Dube, 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. Pankratz, N., Schick, U.M., Zhou, Y., Zhou, W., Ahluwalia, T.S., Allende, M.L., Auer, P.L., Bork-Jensen, J., Brody, J.A., Chen, M.H., et al. (In Press). Meta-analysis of rare and common exome chip variants identifies S1PR4 and other novel genes influencing blood cell traits.

14. Nalls, M.A., Wilson, J.G., Patterson, N.J., Tandon, A., Zmuda, J.M., Huntsman, S., Garcia, M., Hu, D., Li, R., Beamer, B.A., et al. (2008). Admixture mapping of white cell count:

genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. Am J Hum Genet 82, 81-87.

15. 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.

16. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Magi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E., et al. (2014). Quality control and conduct of genome-wide association meta-analyses. Nat Protoc 9, 1192-1212.

17. 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.

18. 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.

19. 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.

20. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190-2191.

21. 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.

22. 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.

23. 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. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 47, 1121-1130.

24. 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. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 47, 979-986.

25. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376-381.

26. Bentham, J., Morris, D.L., Cunninghame Graham, D.S., Pinder, C.L., Tombleson, P., Behrens, T.W., Martin, 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.

27. International Multiple Sclerosis Genetics, C., Wellcome Trust Case Control, C., Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., et al. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476, 214-219.

28. 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. (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat Genet 44, 1341-1348.

29. 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.

30. Dubois, P.C., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A., Adany, R., Aromaa, A., et al. (2010). Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 42, 295-302.

31. 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. (2015). International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. Nat Commun 6, 8019.

32. 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, i185-194.

33. 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. (in press). Several new pleiotropic variants associated with red blood cell traits identified by exome genotyping. Am J Hum Genet.

34. 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. (in press). Platelet-related variants identified by exome chip meta-analysis in 157,293 individuals. Am J Hum Genet.

35. 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.

36. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164.

37. 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.

38. 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. (2013). Genetic variation associated with circulating monocyte count in the eMERGE Network. Hum Mol Genet 22, 2119-2127.

39. Global Lipids Genetics, C., Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., et al. (2013). Discovery and refinement of loci associated with lipid levels. Nat Genet 45, 1274-1283.

40. 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.

41. 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.

42. 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.

43. Nykjaer, A., Moller, 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.

44. 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.

45. Shiow, L.R., Rosen, D.B., Brdickova, 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.

46. 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.

47. Zannettino, A.C., Buhring, 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.

48. 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.

49. 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, D930-934.

50. Gelsi-Boyer, V., Trouplin, V., Adelaide, 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

56. Yanez, 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.

57. Holtschke, T., Lohler, 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.

58. Fotiadis, D., Kanai, Y., and Palacin, M. (2013). The SLC3 and SLC7 families of amino acid transporters. Mol Aspects Med 34, 139-158.

59. 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 U S A 94, 13187-13192.

60. 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.

61. 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.

62. 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.

63. Elmore, S. (2007). Apoptosis: a review of programmed cell death. Toxicol Pathol 35, 495-516.

64. 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.

65. 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 U S A 111, 14846-14851.

66. 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.

67. Chiquet-Ehrismann, R., and Tucker, R.P. (2011). Tenascins and the importance of adhesion modulation. Cold Spring Harb Perspect Biol 3.

68. 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.

69. 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.

70. 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.

71. 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.

72. Utsugi, M., Dobashi, K., Ono, A., Ishizuka, T., Matsuzaki, S., Hisada, T., Shimizu, Y., Kawata, T., Aoki, H., Kamide, Y., et al. (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.

73. 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.

74. 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.

75. 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.

76. Orru, 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.

77. Sancho, D., Gomez, M., Viedma, F., Esplugues, E., Gordon-Alonso, M., Garcia-Lopez, M.A., de la Fuente, H., Martinez, A.C., Lauzurica, P., and Sanchez-Madrid, F. (2003). CD69 downregulates autoimmune reactivity through active transforming growth factor-beta production in collagen-induced arthritis. J Clin Invest 112, 872-882.

78. 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.

79. 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.

80. 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.

81. Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., Consortium, U.K.B.E., North American Brain Expression, C., Coin, L., et al. (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 17, 1418-1428.

82. 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.

83. 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.

84. Cserti, C.M., and Dzik, W.H. (2007). The ABO blood group system and Plasmodium falciparum malaria. Blood 110, 2250-2258.

85. 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.

86. Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schutte, 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.

87. 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.

88. 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.

89. Selleri, C., Montuori, N., Ricci, P., Visconte, V., Carriero, M.V., Sidenius, N., Serio, B., Blasi, F., Rotoli, B., Rossi, G., et al. (2005). Involvement of the urokinase-type plasminogen activator receptor in hematopoietic stem cell mobilization. Blood 105, 2198-2205.

90. 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.

91. 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.

92. 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.

93. 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.

94. 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.

95. 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.

96. 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. (2013). Genome wide association analysis of a founder population identified TAF3 as a gene for MCHC in humans. PLoS One 8, e69206.

97. 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.

Figure Titles and Legends

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.

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). AD, Alzheimer's disease; AS, Ankylosing spondylitis; BAS, basophils; CD, Crohn's 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.

Figure S1. Quantile-quantile plots of p-values of white blood cell traits. Results from 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 S2. Manhattan plots of discovery p-values of white blood cell traits in African Americans.

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

Table 2. Variants associated with white blood cell traits

									Discovery			Replication				Combined Meta-analysis				
Trait (population)	dbSNPID	Chr	Pos	Alt/Ref	EAF	Gene	Annotation	AA Substitution	N	Beta (SE)	P	N	EAF	Beta (SE)	P	N	EAF	Beta (SE)	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)	4.01E-10	17,897	0.6	-0.044 (0.012)	2.66E-04	126,493	0.58	-0.032 (0.004)	8.13E-13	0.28
WBC (All)	rs2276853	3	47,282,303	A/G	0.58	*KIF9*	Missense	p.Arg573Trp	132,764	0.023 (0.004)	3.65E-08	17,897	0.6	0.025 (0.012)	3.00E-02	150,661	0.58	0.023 (0.004)	3.29E-09	0.86
WBC (All)	rs185819[a]	6	32,050,067	C/T	0.51	*TNXB*	Missense	p.His1161Arg	132,764	0.031 (0.005)	4.02E-10	17,897	0.47	0.034 (0.015)	2.24E-02	150,661	0.51	0.031 (0.005)	2.85E-11	0.83
WBC (All)	rs9374080	6	109,616,420	C/T	0.43	*CCDC162P*	Intronic regulatory		132,764	0.023 (0.004)	4.01E-08	17,897	0.46	0.025 (0.011)	2.55E-02	150,661	0.43	0.023 (0.004)	3.15E-09	0.84
WBC (EA)	rs3747869	10	73,520,632	C/A	0.9	*C10orf54 (DD1α)*	Missense	p.Asp187Glu	108,596	0.040 (0.007)	4.26E-08	17,897	0.9	0.083 (0.018)	6.40E-06	126,493	0.9	0.046 (0.007)	1.42E-11	0.03
WBC (EA)	rs1935	10	64,927,823	G/C	0.49	*JMJD1C*	Missense	p.Glu2353Asp	108,596	-0.026 (0.005)	2.46E-08	17,897	0.46	-0.027 (0.012)	2.06E-02	126,493	0.49	-0.026 (0.004)	1.57E-09	0.93
WBC (EA)	rs1292053	17	57,963,537	G/A	0.45	*TUBD1*	Missense	p.Met76Thr	108,596	-0.03 (0.004)	1.28E-11	17,897	0.44	-0.027 (0.011)	1.51E-02	126,493	0.45	-0.030 (0.004)	6.55E-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)	2.51E-10	17,897	0.15	-0.052 (0.015)	7.13E-04	103,582	0.16	-0.044 (0.006)	8.34E-13	0.6
WBC (EA)	rs3865444	19	51,727,962	A/C	0.31	*CD33*	Upstream		86,936	-0.037 (0.005)	3.51E-12	17,897	0.32	-0.033 (0.012)	5.14E-03	104,833	0.31	-0.036 (0.005)	6.81E-14	0.77
WBC (All)	rs2836878	21	40,465,534	A/G	0.26	*ETS2-PSMG1*	Intergenic		132,764	-0.025 (0.005)	8.36E-08	17,897	0.26	-0.026 (0.012)	3.44E-02	150,661	0.26	-0.025 (0.004)	8.41E-09	0.89
Neu (EA)	rs3747869	10	73,520,632	C/A	0.9	*C10orf54 (DD1α)*	Missense	p.Asp187Glu	60,851	0.053 (0.010)	2.11E-08	16,669	0.9	0.073 (0.019)	1.17E-04	77,520	0.9	0.057 (0.009)	1.65E-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)	1.54E-08	16,669	0.15	-0.044 (0.016)	5.55E-03	72,781	0.16	-0.046 (0.007)	3.01E-10	0.87
Mon (All)	rs4917014	7	50,305,863	G/T	0.28	*C7orf72-IKZF1*	Intergenic		57,183	-0.038 (0.007)	1.97E-08	16,669	0.32	-0.048 (0.012)	8.92E-05	73,852	0.29	-0.040 (0.006)	9.75E-12	0.48
Mon (EA)	rs11625112	14	23,596,740	G/A	0.46	*SLC7A8*	Intronic		44,325	-0.038 (0.007)	3.82E-08	16,669	0.45	-0.031 (0.012)	7.04E-03	60,994	0.46	-0.036 (0.006)	1.03E-09	0.62
Mon (EA)	rs11642873[a]	16	85,991,705	C/A	0.2	*IRF8-LINC01082*	Intergenic		44,325	0.057 (0.008)	1.41E-11	16,669	0.2	0.113 (0.014)	6.17E-15	60,994	0.2	0.072 (0.007)	1.40E-22	0.001
Mon (All)	rs1292053	17	57,963,537	G/A	0.45	*TUBD1*	Missense	p.Met76Thr	57,183	-0.036 (0.006)	2.55E-09	16,669	0.44	-0.040 (0.012)	6.08E-04	73,852	0.45	-0.037 (0.005)	6.53E-12	0.76
Lym (EA)	rs2229094	6	31,540,556	C/T	0.26	*LTA*	Missense	p.Cys13Arg	47,105	0.046 (0.008)	1.89E-08	16,711	0.25	0.078 (0.018)	8.54E-06	63,816	0.26	0.051 (0.007)	3.14E-12	0.09
Lym (EA)	rs4763879	12	9,910,164	A/G	0.36	*CD69*	Intronic regulatory		47,105	-0.038 (0.007)	3.08E-08	16,711	0.36	-0.038 (0.012)	1.36E-03	63,816	0.36	-0.038 (0.006)	1.59E-10	0.99
Bas (EA)	rs2295764	20	31,025,163	G/A	0.36	*ASXL1*	3' UTR		44,138	-0.042 (0.007)	3.28E-09	15,770	0.36	-0.031 (0.012)	9.78E-03	59,908	0.36	-0.039 (0.006)	1.46E-10	0.43

Abbreviations: 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.

Table 3. Association of white blood cell trait variants with immune-mediated diseases and clinical phenotypes in previous genome-wide association studies

Trait (population)	dbSNPID	Chr	Pos	Alt/Ref	Gene	Phenotype	Sample size	P[a]
WBC(EA)	rs1260328	2	27,730,940	C/T	*GCKR*	Inflammatory bowel disease	96,486	1.27E-04
Lym(EA)	rs2229094[b]	6	31,540,556	C/T	*LTA*	Crohn's disease	69,268	7.81E-07
Lym(EA)	rs2229094[b]	6	31,540,556	C/T	*LTA*	Systemic lupus erythematosus	23,209	3.09E-07
Lym(EA)	rs2229094[b]	6	31,540,556	C/T	*LTA*	Primary biliary cirrhosis	21,216	1.31E-05
WBC(All)	rs185819	6	32,050,067	C/T	*TNXB*	Multiple sclerosis[c]	22,850	2.16E-06
WBC(All)	rs185819	6	32,050,067	C/T	*TNXB*	Type 1 diabetes	33,394	3.29E-09
WBC(All)	rs185819	6	32,050,067	C/T	*TNXB*	Ulcerative colitis	72,647	2.91E-06
WBC(All)	rs185819	6	32,050,067	C/T	*TNXB*	Systemic lupus erythematosus[c]	23,209	2.32E-37
WBC(All)	rs185819	6	32,050,067	C/T	*TNXB*	Rheumatoid arthritis	103,558	3.90E-53
Mon(All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	Systemic lupus erythematosus	32,444	8.10E-05
Mon(All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	Inflammatory bowel disease	96,486	4.59E-05
Mon(All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	Crohn's disease	69,268	1.49E-04
Mon(All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	Stevens-Johnson syndrome/toxic epidermal necrolysis	1,129	8.00E-11
Mon(All)	rs4917014	7	50,305,863	G/T	*C7orf72-IKZF1*	Selective immunoglobulin A deficiency	2,748	2.80E-23
Lym(EA)	rs4763879	12	9,910,164	A/G	*CD69*	Type 1 diabetes	38,522	1.90E-11
Lym(EA)	rs4763879	12	9,910,164	A/G	*CD69*	Multiple sclerosis	38,135	2.18E-05
Lym(EA)	rs4763879	12	9,910,164	A/G	*CD69*	Selective immunoglobulin A deficiency	2,748	1.90E-11
Mon(EA)	rs11642873	16	85,991,705	C/A	*IRF8-LINC01082*	Systemic lupus erythematosus	23,209	3.56E-10
Mon(EA)	rs11642873	16	85,991,705	C/A	*IRF8-LINC01082*	Systemic sclerosis	14,853	2.30E-12
WBC(EA); Mon(All)	rs1292053	17	57,963,537	G/A	*TUBD1*	Multiple sclerosis	38,135	7.47E-06
WBC(EA); Mon(All)	rs1292053	17	57,963,537	G/A	*TUBD1*	Crohn's disease	96,486	8.53E-06
WBC(EA); Mon(All)	rs1292053	17	57,963,537	G/A	*TUBD1*	Inflammatory bowel disease	96,486	9.61E-05
WBC(EA); Neu(EA)	rs4760	19	44,153,100	G/A	*CD87 (PLAUR)*	Ulcerative colitis	72,647	1.51E-04
WBC(EA)	rs3865444	19	51,727,962	A/C	*CD33*	Alzheimer's disease	59,716	1.60E-09
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	4.90E-12
WBC(All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	Crohn's disease	69,268	2.43E-06
WBC(All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	Ulcerative colitis	72,647	2.05E-20
WBC(All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	Inflammatory bowel disease	96,486	3.70E-22
WBC(All)	rs2836878	21	40,465,534	A/G	*ETS2-PSMG1*	Selective immunoglobulin A deficiency	2,748	1.40E-08

[a] Significant results are shown after correcting for multiple testing of 16 variants and 15 diseases (p-value < 2.08E-04). 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.

Abbreviations: 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.