Human Molecular Genetics, 2014, Vol. 23, No. 25 **6944–6960**
doi:10.1093/hmg/ddu401
Advance Access published on August 5, 2014

Trans-ethnic meta-analysis of white blood cell phenotypes

Margaux F. Keller[1,7,†], Alexander P. Reiner[8,12,†], Yukinori Okada[13,15,†], Frank J.A. van Rooij[16,19,†], Andrew D. Johnson[20,22], Ming-Huei Chen[21,22], Albert V. Smith[23,24], Andrew P. Morris[25,26], Toshiko Tanaka[3], Luigi Ferrucci[3], Alan B. Zonderman[4], Guillaume Lettre[27,28], Tamara Harris[2], Melissa Garcia[2], Stefania Bandinelli[29], Rehan Qayyum[30], Lisa R. Yanek[30], Diane M. Becker[30], Lewis C. Becker[30,31], Charles Kooperberg[12], Brendan Keating[32,33], Jared Reis[34], Hua Tang[35], Eric Boerwinkle[36], Yoichiro Kamatani[13], Koichi Matsuda[37], Naoyuki Kamatani[13], Yusuke Nakamura[37,38,39], Michiaki Kubo[14], Simin Liu[40,41], Abbas Dehghan[16,19], Janine F. Felix[16,19], Albert Hofman[16,19], André G. Uitterlinden[16,18,19], Cornelia M. van Duijn[16,19], Oscar H. Franco[16,17,19], Dan L. Longo[5], Andrew B. Singleton[1], Bruce M. Psaty[9,10,11,42], Michelle K. Evans[6], L. Adrienne Cupples[22,43], Jerome I. Rotter[44,45], Christopher J. O'Donnell[20,22], Atsushi Takahashi[13,‡], James G. Wilson[46,‡], Santhi K. Ganesh[47,48,‡,*] and Mike A. Nalls[1,‡,*] for the CHARGE Hematology, COGENT, and BioBank Japan Project (RIKEN) Working Groups[¶]

[1]Laboratory of Neurogenetics, [2]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA, [3]Longitudinal Studies Section, Clinical Research Branch, [4]Behavioral Epidemiology Section, Laboratory of Epidemiology & Population Sciences, National Institute on Aging Intramural Research Program, National Institutes of Health, Baltimore, MD, USA, [5]Laboratory of Genetics, [6]Health Disparities Research Section, Clinical Research Branch, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA, [7]Department of Biological Anthropology, Temple University, Philadelphia, PA, USA, [8]Department of Epidemiology, [9]Cardiovascular Health Research Unit, [10]Department of Medicine, [11]Department of Epidemiology and Health Services, University of Washington, Seattle, WA, USA, [12]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [13]Laboratory for Statistical Analysis, [14]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, [15]Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan, [16]Department of Epidemiology, [17]ErasmusAGE, Department of Epidemiology, [18]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands, [19]Consortium for Healthy Aging (NGI-NCHA), The Netherlands Genomics Initiative, Leiden, The Netherlands, [20]Cardiovascular Epidemiology and Human Genomics Branch, NHLBI Division of Intramural Research, Bethesda, MD, USA, [21]Department of Neurology, Boston University School of Medicine, Boston, MA, USA, [22]NHLBI Framingham Heart Study, Bethesda, MD, USA, [23]Icelandic Heart Association, Kopavogur, Iceland, [24]University of Iceland, Reykjavik, Iceland, [25]Genetic and Genomic Epidemiology Unit, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK, [26]Department of Biostatistics, University of Liverpool, Liverpool, UK, [27]Montreal Heart Institute, Montréal, Canada, [28]Département de Médecine, Université de Montréal, Montréal, Canada, [29]Geriatric Rehabilitation Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy, [30]GeneSTAR Research Program, Division of General Internal Medicine , [31]Division of Cardiology, Johns Hopkins School of Medicine, Baltimore, MD, USA, [32]Center for Applied Genomics, Children's Hospital of Philadelphia, PA, USA, [33]Department of Pediatrics, University of Pennsylvania,

[*]To whom correspondence should be addressed at: Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA. Tel: +1 7347644500; Fax: +1 7349368266; Email: sganesh@med.umich.edu; Laboratory of Neurogenetics, Building 35, 1A-1014, 35 Convent Drive Bethesda, MD 20892, USA. Tel: 301 451 3831; E-mail: nallsm@mail.nih.gov
[†]These authors contributed equally to this work.
[‡]These authors contributed equally to this work.
[¶]A full list of collaborators from CHARGE, COGENT and RIKEN can be found in Supplementary Material.

PA, USA, [34]Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, Bethesda, MD, USA, [35]Stanford University School of Medicine, Stanford, CA 94305, USA, [36]The Brown Foundation, Institute of Molecular Medicine for the Prevention of Human Diseases, University of Texas, Houston, TX, USA, [37]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [38]Department of Medicine, [39]Department of Surgery, Center for Personalized Therapeutics, The University of Chicago, Chicago, IL, USA, [40]Department of Epidemiology, [41]Department of Medicine, Brown University, Providence, RI, USA, [42]Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA, [43]Boston University Department of Statistics, Boston, MA, USA, [44]Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA, [45]Division of Genetic Outcomes, Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, USA, [46]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA [47]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA and [48]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA

Received April 25, 2014; Revised and Accepted July 31, 2014

White blood cell (WBC) count is a common clinical measure used as a predictor of certain aspects of human health, including immunity and infection status. WBC count is also a complex trait that varies among individuals and ancestry groups. Differences in linkage disequilibrium structure and heterogeneity in allelic effects are expected to play a role in the associations observed between populations. Prior genome-wide association study (GWAS) meta-analyses have identified genomic loci associated with WBC and its subtypes, but much of the heritability of these phenotypes remains unexplained. Using GWAS summary statistics for over 50 000 individuals from three diverse populations (Japanese, African-American and European ancestry), a Bayesian model methodology was employed to account for heterogeneity between ancestry groups. This approach was used to perform a trans-ethnic meta-analysis of total WBC, neutrophil and monocyte counts. Ten previously known associations were replicated and six new loci were identified, including several regions harboring genes related to inflammation and immune cell function. Ninety-five percent credible interval regions were calculated to narrow the association signals and fine-map the putatively causal variants within loci. Finally, a conditional analysis was performed on the most significant SNPs identified by the trans-ethnic meta-analysis (MA), and nine secondary signals within loci previously associated with WBC or its subtypes were identified. This work illustrates the potential of trans-ethnic analysis and ascribes a critical role to multi-ethnic cohorts and consortia in exploring complex phenotypes with respect to variants that lie outside the European-biased GWAS pool.

INTRODUCTION

White blood cells (WBCs) are critically involved in the body's immune system, serving as a primary defense mechanism against foreign pathogens. WBC count is used as a clinical marker of inflammation status, and higher WBC count has been associated with a risk of cardiovascular disease, cancer mortality and all-cause mortality (1–3). Elevated WBC count is also associated with disease risk factors including increasing age, high blood pressure, cigarette smoking, adiposity and increasing plasma inflammatory markers (4).

WBCs are classified into five subtypes according to their morphology and functions, including neutrophils, basophils, eosinophils, lymphocytes and monocytes. Total WBC count is highly variable even among healthy individuals of the same population (5). WBC count is a moderately heritable phenotype, with h^2 estimates ranging from 0.14 to 0.40 across the WBC subtypes (6). Additionally, between 25 and 50% of individuals of African descent exhibit benign ethnic neutropenia, characterized by low neutrophil counts, due to a regulatory variant in the Duffy antigen receptor for chemokines (*DARC*) gene (5,7,8). Given the importance of WBC in both host defense and, potentially, pathologic inflammation, elucidation of additional genetic mechanisms responsible for regulating while blood cell count could have a substantial medical impact.

Admixture mapping and genome-wide association studies (GWAS) performed on cohorts of differing continental ancestry, including European, Japanese and African-American, have been successful in identifying multiple loci associated with WBC phenotypes (5,8–14). The joint effects of these loci generally explain only a small portion of the overall heritability of either total WBC or WBC subtypes. Some prior GWAS have not defined the subtypes of WBC that are driving their observed associations; however, neutrophils in particular are often implicated (9,11). Furthermore, the loci identified by GWAS generally encompass large genomic regions, often containing many genes and variants with comparable association signals. Thus, fine-mapping methods aimed at pinpointing association signals more precisely are needed (15).

Recently, the 1000 Genomes Project and phase three of the HapMap project released comprehensive reference panels for a number of ethnic groups, including African, Asian and

additional European populations (16,17). Imputation using these higher density reference panels allows inference of genotypes not captured by genotyping arrays, markedly increasing the breadth of genetic variation that can be included in association tests. This has provided new opportunities both to detect novel loci and to refine the localization of association signals for a number of phenotypes, including WBC phenotypes.

Trans-ethnic meta-analysis (MA) potentially offers a more comprehensive view of the genetic variation that is associated with a trait, but traditional fixed-effects MA methods do not adequately address heterogeneity in allelic effects, allele frequencies or differences in linkage disequilibrium between ethnicities (18). For example, in a previous fixed-effects analysis of the cohorts included in the current study, only 152 of 161 single nucleotide polymorphisms (SNPs) that had been associated with WBC phenotypes in earlier analyses were replicated at a Bonferroni-corrected significance threshold of $P < 3.57E-3$, and no novel associations were observed (9). Whereas random-effects-based methods of MA do account for inter-study heterogeneity, they lose statistical power in the setting of high levels of heterogeneity that may result from experimental or statistical differences in study design (15,19).

These shortcomings have been addressed in the software package, MANTRA (Meta-Analysis of Trans-ethnic Association studies), which allows for heterogeneity between diverse ethnic groups and provides increased power and mapping resolution compared with random-effects-based methods (15). In the current study, we used MANTRA to combine summary results of ancestry-specific GWAS of WBC traits in three distinct populations. We identify novel loci associated with WBC count, assess heterogeneity in allelic effects between ancestry groups and improve fine-mapping resolution of some previously identified regions.

RESULTS

Descriptive statistics for each cohort are found in Table 1. In the trans-ethnic MANTRA analyses, we observed strong evidence of association, defined by a $\log_{10}$ Bayes factor (BF) of >6, at 10 previously identified loci and six novel loci, and detected nine secondary signals within 500 kb of a previously identified locus. The population-specific and trans-ethnic results for the established and novel loci associated with each WBC phenotype (total WBC, neutrophil and monocyte count) are summarized in Table 2. Cohort-level Manhattan plots are shown in Figure 1. Of the 15 previously identified variants (at 10 loci), six of the six monocyte associations, two of the four neutrophil associations and two of the five WBC count associations were initially identified in the original GWAS papers from which the data employed by this analysis are drawn.

Regions previously identified by single-ethnicity GWAS reappeared in the MANTRA trans-ethnic analyses, but in some instances, the index SNP from the original publication was not the most significant. These include rs4065321 and rs17609240 on 17q21.1 (WBC count and neutrophil count), rs2517524 on 6p21.33 (WBC count) and rs10956483 on 8q24.21 (monocyte count). Additionally, rs2814778, a marker identifying the Duffy null blood group antigen and located on the *DARC* gene of chromosome 1 at position (b37) 159 174 683, was available only in the COGENT data. This marker accounts for 20% of population variance in the WBC of African ancestry populations and is monomorphic in non-African populations.

Novel associations

In addition to replicating known variants, the trans-ethnic analysis identified six novel trait–locus associations (Table 2). For neutrophil counts, novel findings include rs6936204 in region 6p21.32, located nearest to *AK123889*. This region is very near known locus 6p21.33, which was previously associated with WBC and lymphocyte counts (9). These loci are near the *HLA* region; thus, it is possible that population stratification is driving this association (20). These variants are not in linkage disequilibrium (LD) with any of the known *HLA* markers, but as meiotic crossovers are known to cluster around *HLA*, it is possible that these variants are separated from this region by a recombination hotspot. Novel association for WBC count includes rs10932765 in region 2q35, located near *ARPC2*, which has previously been associated with monocyte count and inflammatory bowel disease (12,21). This region is notable because of its proximity to *IL8RA*, which encodes *CXCR1*. *CXCR1*, the receptor for the chemokine *IL-8*, is a mediator of inflammatory responses; interestingly, the Duffy antigen

Table 1. Descriptive statistics

Study Population ancestry	RIKEN Japanese	CHARGE European	COGENT African-American
Total WBC			
WBC count: mean (SD) in cells $\times 10^3$/ml	6.20 (1.67), $n = 16\,843$	5.72 (1.24), $n = 19\,509$	5.94 (1.88), $n \sim 16\,388$
Neutrophils: mean (SD) in cells $\times 10^3$/ml	3.80 (1.40), $n = 9802$	3.52 (1.06), $n = 16\,550$	3.57 (1.57), $n \sim 7391$
Monocytes: mean (SD) in cells $\times 10^3$/ml	0.36 (0.22), $n = 9810$	0.43 (0.14), $n = 16\,550$	0.36 (0.17), $n \sim 7369$
Covariates			
Age in years: mean (SD)	63.5 (10.6)	63.8 (8.9)	50.9 (7.6)
BMI: mean (SD) kg/m^2	23.1 (3.6)	NA	NA
% Female	39.1	53.5	65.6
% Current smoker	52.6	17.6	NA
Sample size			
Total *N*	17 218	19 509	16 388

Counts are reported in thousands of cells per milliliter of blood (cells $\times 10^9$/l).
SD, standard deviation; NA, not available.

Table 2. Loci defined in MANTRA analysis

Lead SNP	Chr	base pair (BP) (bd 37)	Effect	Other	MANTRA trans-ethnic meta-analysis				European ancestry GWAS meta-analysis						Japanese ancestry GWAS meta-analysis						African-American ancestry GWAS meta-analysis					
					$\log_{10}$ BF (bayes factor) assocation	PPA	Sample size	Gene(s)	EAF	Beta	SE	*P*-value	Sample size	R^{2a}	EAF	Beta	SE	*P*-value	Sample size	R^2	EAF	Beta	SE	*P*-value	Sample size	R^{2a}
WBC count—established loci																										
rs2518564*	1	159 062 436	A	G	34 1.59	1	51 768	DARC	0.293	0.007	0.003	2.81E−02	19 509	0.883	0.7159	0	0.003	0.9936	16 843	0.991	0.1725	0.1905	0.0048	1.39E−149	15 416	1
rs1371799	4	74 977 837	C	T	13.5	1	52 740	CXCL2	0.572	−0.012	0.003	2.14E−05	19 509	0.985	0.461	−0.012	0.003	3.63E−05	16 843	0.921	0.761	0.023	0.0041	5.21E−04	16 388	1
rs4895441	6	135 426 573	A	G	9.64	0.165	52 703	Intergenic	0.347	0.011	0.003	3.42E−04	19 509	1	0.628	0.018	0.003	8.99E−11	16 843	1	0.902	0.0002	0.0059	9.90E−01	16 351	1
rs445	7	92 408 370	C	T	16.8	1	52 740	CDK6	0.776	0.019	0.005	4.51E−04	19 509	1	0.685	0.019	0.003	8.99E−11	16 843	1	0.193	−0.0249	0.0049	9.90E−01	16 388	1
rs4794822	17	38 156 712	C	T	29.64	1	52 740	PSMD3	0.586	−0.028	0.003	3.23E−23	19 509	0.999	0.481	−0.019	0.003	2.90E−12	16 843	0.999	0.325	0.0128	0.0038	4.33E−02	16 388	0.918
WBC count—novel loci																										
rs10932765	2	219 099 484	C	T	6.95	0.025	52 740	ARPC2	0.503	0.011	0.003	3.36E−05	19 509	1	0.58	0.012	0.003	1.69E−05	16 843	0.986	0.877	0.0039	0.0056	7.26E−01	16 388	1
rs6734238	2	113 841 030	A	G	6.62	0.028	52 691	IL1F10	0.451	−0.009	0.003	1.46E−03	19 509	1	0.967	−0.018	0.008	2.49E−02	16 843	0.907	0.553	−0.0158	0.0035	5.92E−03	16 339	1
rs2853946	6	31 247 203	A	T	12.16	0.009	52 740	HLA-B	0.348	−0.017	0.003	1.28E−08	19 509	1	0.717	−0.013	0.003	8.19E−06	16 843	0.976	0.512	−0.0106	0.0035	7.65E−02	16 388	1
rs2163950	8	130 597 585	A	C	7.75	0.065	52 740	Intergenic	0.054	−0.017	0.006	5.91E−03	19 509	0.861	0.174	−0.021	0.004	5.36E−09	16 843	0.995	0.124	−0.0053	0.0058	6.44E−01	16 388	0.788
Monocyte count—established loci																										
rs1449263	2	182 319 301	C	T	19.24	1	33 729	ITGA4	0.541	−0.036	0.005	6.70E−14	16 550	0.942	0.626	−0.033	0.006	8.32E−08	9,810	0.994	0.58	0.0087	0.0031	4.68E−03	7369	0.965
rs12988934	2	182 323 665	C	T	7.75	0.022	26 360	ITGA4	0.789	−0.034	0.01	1.01E−03	16 550	0.511	0.732	−0.041	0.008	1.42E−07	9810	0.73	0.018	−0.0165	0.0854	8.47E−01	1803	0.418
rs9880192	3	128 297 569	C	G	8.3	1	33 745	C3orf27	0.441	−0.028	0.005	1.35E−08	16 550	0.867	0.091	−0.037	0.012	1.21E−03	9810	0.804	0.188	0.0037	0.0047	4.31E−01	7385	0.674
rs3095254	6	31 221 668	C	G	6.81	1	33 745	MHC	0.377	0.008	0.005	1.19E−01	16 550	0.976	0.461	0.035	0.006	8.27E−09	9810	0.896	0.562	0.0031	0.0031	3.31E−01	7385	0.935
rs1991866	8	130 624 105	C	G	15.3	1	33 598	Intergenic	0.451	−0.032	0.005	4.58E−11	16 550	0.96	0.598	−0.034	0.006	7.38E−08	9810	0.892	0.52	−0.0046	0.0031	1.32E−01	7238	1
rs10980800	9	113 915 905	C	T	11.63	1	33 692	Intergenic	0.769	0.044	0.006	1.10E−14	16 550	0.988	0.056	0.004	0.013	7.61E−01	9810	0.941	0.787	−0.0049	0.0037	1.85E−01	7332	1
Monocyte count—novel loci																										
rs2047076	5	76 058 509	C	T	6.03	1	33 729	Intergenic	0.714	0.005	0.006	3.80E−01	16 550	0.999	0.9998	2.678	0.474	1.64E−08	9,810	1	0.037	0.0023	0.0091	7.97E−01	7369	0.789
Neutrophil count—established loci																										
rs7667376	4	74 967 890	C	T	11.13	1	33 743	CXCL2	0.596	0.017	0.004	1.47E−04	16 550	0.972	0.542	0.017	0.005	5.96E−04	9802	1	0.221	−0.0428	0.008	8.50E−04	7391	1
rs445	7	92 408 370	C	T	10.52	1	33 744	CDK6	0.776	0.022	0.009	1.21E−02	16 550	1	0.685	0.024	0.005	8.96E−06	9802	1	0.203	−0.0479	0.0092	1.15E−03	7392	1
rs8078723	17	38 166 879	C	T	28.06	1	33 693	CSF3, MED24	0.626	0.043	0.004	2.84E−23	16 550	0.997	0.519	0.032	0.005	1.39E−10	9802	0.995	0.678	−0.0071	0.0073	6.19E−01	7341	1
rs4794822	17	38 156 712	C	T	28.92	1	33 753	PSMD3	0.586	−0.043	0.004	3.64E−23	16 550	0.999	0.481	−0.032	0.005	7.09E−11	9802	0.999	0.325	0.0048	0.0073	7.41E−01	7401	0.918
Neutrophil count—novel loci																										
rs6936204	6	32 217 092	C	T	6.45	0.048	33 706	AK123889	0.585	0.02	0.004	5.83E−06	16 550	0.991	0.912	0.032	0.009	3.97E−04	9802	0.983	0.279	0.01	0.0074	4.68E−01	7354	1

Strong evidence for association is defined as a BF of >6. Differences in effect alleles between cohorts were corrected before the MANTRA analysis.
PPA, posterior probability of association; SE, standard error; EAF, effect allele frequency.
[a]Consortia summary of imputation qualtiy score.
*Proxy for rs2814778.

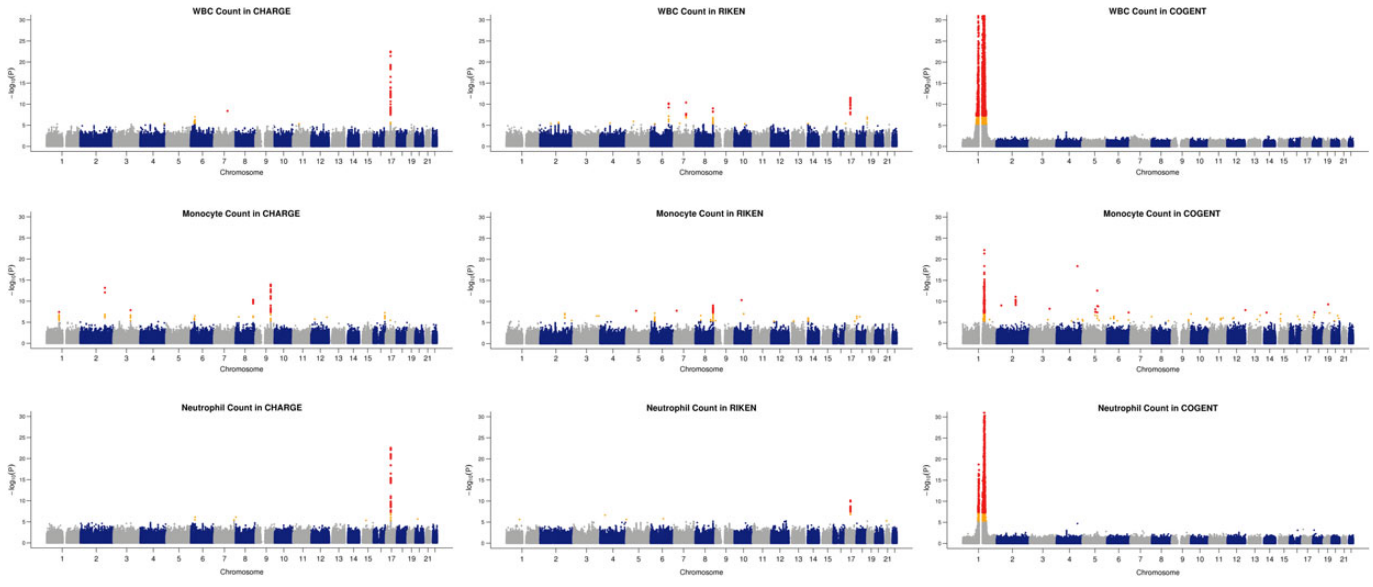

Figure 1. Manhattan plots subset by WBC subtype and cohort. Horizontal axis indicates the chromosomal position. Vertical axis for the blue Manhattan plots indicates $-\log_{10} P$-values from fixed-effects meta-analysis, and vertical axis for the black Manhattan plots indicates BFs from trans-ethnic meta-analysis. Significance values are truncated at 30 on the *y*-axis for clarity and scaling of images. A 50 Mb region of apparent significance surrounds the centromere of chromosome 1 and suggests two distinct peaks; however, this results from the lack of genotyped or imputed SNPs in the region and is a spatial inflation of a truly causal variant in the nearby *DARC* gene [Nalls *et al*. (9) and Reiner *et al*. (12)].

is also a receptor for *IL-8*. Additional novel associations for WBC include rs2163950 at 8q24.21, located in an intergenic region, and rs6734238 in region 2q13, located near *IL1F10* and *IL1RN*. Top SNP rs6734238 tags the *IL-1* gene family locus near 2q13 and has also been associated with C-reactive protein levels in a European ancestry population (22). Notably, a large region of the implicated chromosome 6 associations for WBC, tagged by rs2853946, contains an apparently bimodal signal of association, between 310.00 and 313.00 Mb (Fig. 2). A single novel association for monocyte count was identified by rs2047076 in region 5q13.3, also within an intergenic region of the genome. This region falls between *F2R*, associated with platelet count (23), and *F2RL1* which encodes PAR-2, a monocyte receptor (24).

Variants showing strong evidence for association were examined for heterogeneity in their allelic effects across ancestries, indicated by a posterior probability of heterogeneity (PPH) of >0.5 (Table 3). For the novel monocyte-associated locus on chromosome 5, the RIKEN cohort has a large posterior mean allelic effect (PMAE), whereas the COGENT and CHARGE cohorts have PMAE that are much smaller. This suggests that the association may be specific to the Japanese population, or that the variant tagging this region, rs2047076, may not be a good proxy among European or African-American individuals because of differing LD structure between these populations. The associations reported on chromosome 4 for both WBC and neutrophils exhibit allelic effects in opposite directions between the ancestry groups, which could reflect multiple risk variants, or differing LD structure (1). The novel associations reported on chromosome 2 for WBC have a posterior probability of association (PPA) of $\sim$0.03, and the PMAE of these variants are similar across the three ancestries ($\sim$0.01). PPA is estimated from the weighted average of the alternative models and accounts for the differences in likelihood and statistical power between tests (25). Locus plots for novel associations are shown in Figures 2–7, and locus plots for known associations are shown in Supplementary Material, Figures S1–S11.

To determine how consistent the MANTRA results are across studies, a random-effects MA was also performed using METAL. The results are largely supportive of the known and novel loci reported here, with the exception of those variants with high levels of heterogeneity between populations, such as the novel variant associated with monocytes and located on chromosome 5, rs2047076. The results of this additional MA are found in Supplementary Material, Table S1.

Cohort-level data were imputed before the trans-ethnic analysis was performed. In instances where a novel variant was both imputed across multiple cohorts and exhibits low allele frequencies within those cohorts, additional replication is needed to validate the associations made here. In particular, these include rs2047076, associated with monocytes, and rs2163950, associated with WBC count (Table 2). Without replication, these novel results should be viewed cautiously. We are optimistic that in time, a similarly sized, ancestry-matched cohort will be available to replicate these analyses. In the meantime, we have provided genome-wide summary statistics for the primary MANTRA analyses of monocyte, neutrophil and WBC count traits in Supplementary Material, Table S2.

Fine mapping

Credible sets were defined to assess the extent to which the trans-ethnic analysis improved fine-mapping resolution of known associations. Credible region summary data for associated loci are presented in Supplementary Material, Table S3.

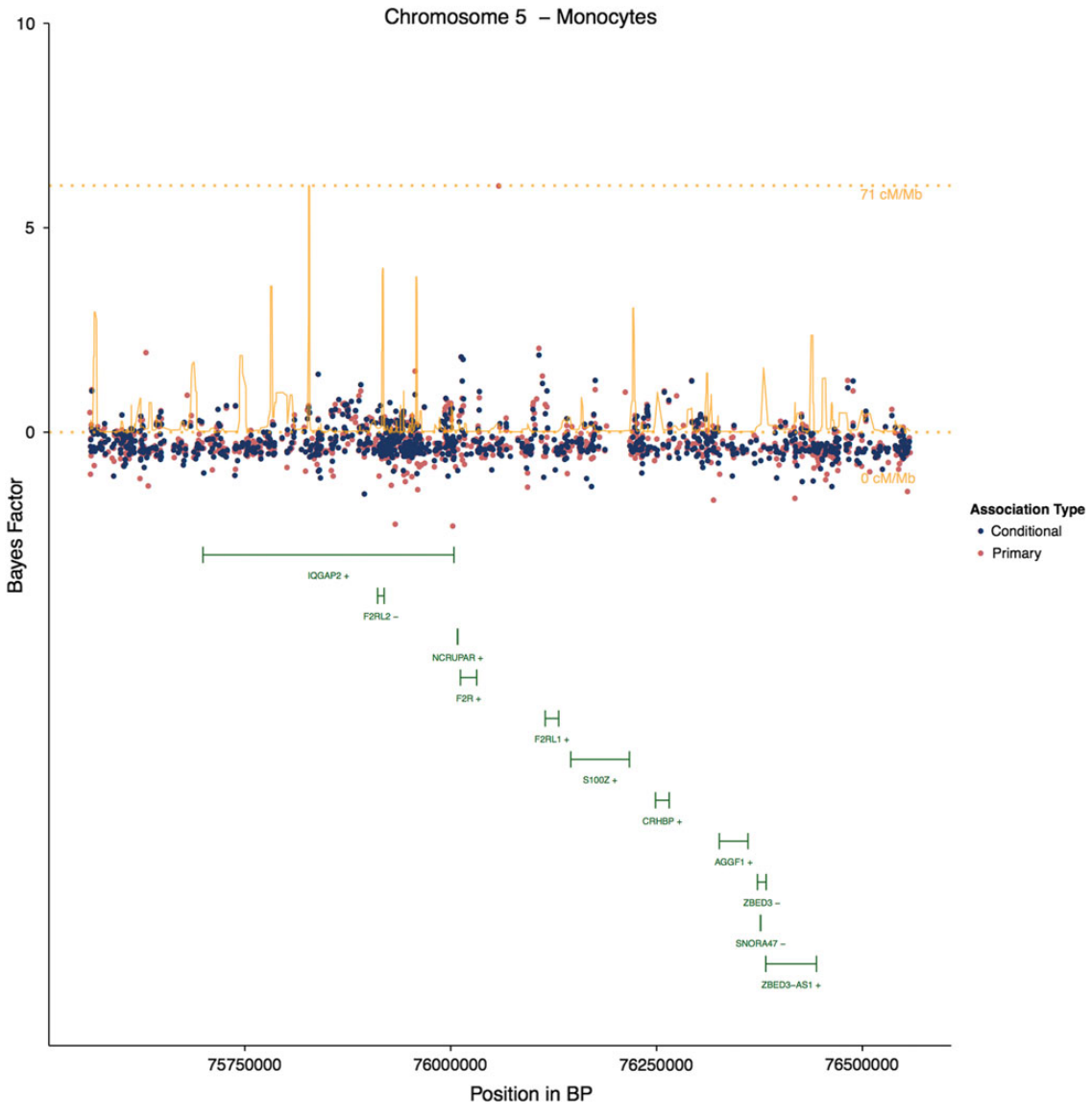

Figure 2. Locus plot for Monocyte association on chromosome 5. Vertical axis indicates BF, and horizontal axes indicate both chromosomal position and gene location.

The major locus affecting WBC levels in African-Americans is located in the *DARC* gene of chromosome 1, spanning 900 000 bp between 158 724 683 and 159 624 683 (3). The main variant associated with this signal is rs2814778, located at position (b37) 159 174 683. Our analyses replicated this finding, identifying a number of significant hits surrounding rs2814778 (chr1: 159 174 683). The surrounding variants are the product of a well-established selective sweep, and proximity to the Duffy null mutations predisposed them to association (26). While rs2814778 was not included in all three cohorts due to the removal of monomorphic SNPs during quality control, credible region analyses of this region identify a single nearby variant, rs2518564 (chr1: 159 062 436), as encompassing 99% of the signal. As the functional variant in the *DARC* locus is present exclusively in the African-American population consortium, we did not expect the this region to show meaningful fine mapping. The use of a proxy variant in high LD with the other two cohorts provides additional evidence of what is already known, this region is highly associated with WBC.

Other previously identified associations with WBC subtypes, however, were substantially narrowed. For example, the 7q21.2 region tagged soley by rs445 in both WBC and neutrophil counts was found to encompass 99% of the association signal for these traits. This variant is located within an intronic section of *CDK6*, a gene in the cyclin-dependent protein kinase family. At another previously identified locus for WBC and neutrophil count, located on chromosome 17, the association signal could be limited with 99% confidence to seven variants across a ∼15 kb region associated with WBC count and two variants across a ∼10 kb region associated with neutrophil count. The previously identified variant tagging this region, rs4794822, is within both of these SNP sets, but individually reaches only 72% confidence in WBC count and 88% confidence in neutrophil count.

Table 3. Heterogeneity of allelic effects

Subtype	Variant details			MANTRA MA		CHARGE		RIKEN		COGENT	
	SNP	Chr	Position (b37)	$\log_{10}$ (BF)	PPH	PMAE	PSD	PMAE	PSD	PMAE	PSD
Novel											
WBC	rs10932765	2	219 099 484	6.95	0.025	0.01068	0.00182	0.01068	0.00182	0.01057	0.00205
WBC	rs6734238	2	113 841 030	6.62	0.028	−0.01193	0.00244	−0.01181	0.00213	−0.01195	0.00227
WBC	rs1371799	4	74 977 837	13.5	1	−0.01164	0.00207	−0.01163	0.00207	0.02273	0.00408
WBC	rs2853946	6	31 247 203	12.16	0.009	−0.01379	0.00179	−0.01379	0.00179	−0.01378	0.00181
WBC	rs2163950	8	130 597 585	7.75	0.065	−0.01688	0.00287	−0.01655	0.00304	−0.01614	0.00346
MONO	rs2047076	5	76 058 509	6.03	1	2.50643	0.47953	0.00458	0.00498	0.00457	0.00495
NEU	rs7667376	4	74 967 890	11.13	1	0.011	0.00199	0.01104	0.00199	−0.02275	0.00405
NEU	rs6936204	6	32 217 092	6.45	0.048	0.01028	0.00356	0.0096	0.00259	0.00681	0.00477
Known											
WBC	rs2518564*	1	159 062 436	341.59	1	0.00318	0.00241	0.00363	0.00258	0.19017	0.00469
WBC	rs4895441	6	135 426 573	9.64	0.165	0.01368	0.00248	0.01282	0.00239	0.01169	0.00446
WBC	rs445	7	92 408 370	16.8	1	0.01905	0.00254	0.01899	0.00297	−0.02476	0.00476
WBC	rs4794822	17	38 156 712	29.64	1	−0.02088	0.00309	−0.02491	0.00337	0.01249	0.00391
MONO	rs1449263	2	182 319 301	19.24	1	−0.03449	0.00399	−0.03484	0.00378	0.00857	0.00316
MONO	rs12988934	2	182 323 665	7.75	0.022	−0.03872	0.00621	−0.03853	0.00627	NA	NA
MONO	rs9880192	3	128 297 569	8.3	1	−0.03049	0.00687	−0.02917	0.0046	0.00315	0.00483
MONO	rs3095254	6	31 221 668	6.81	1	0.03449	0.00622	0.00473	0.00306	0.00439	0.00267
MONO	rs1991866	8	130 624 105	15.3	1	−0.0322	0.00418	−0.03217	0.00392	−0.00466	0.00316
MONO	rs10980800	9	113 915 905	11.63	1	0.0087	0.01581	0.043	0.00581	−0.00475	0.00379
NEU	rs445	7	92 408 370	10.52	1	0.01905	0.00254	0.01899	0.00297	−0.02476	0.00476
NEU	rs8078723	17	38 166 879	28.06	1	0.01851	0.00278	NA	NA	−0.01105	0.00378
NEU	rs4794822	17	38 156 712	28.92	1	−0.02088	0.00309	−0.02491	0.00337	0.01249	0.00391

Variants encompassing the 95% credible region of an associated region, as identified using the top hits from the MANTRA analysis, are presented for each subtype. Ancestry-specific posterior mean allelic effects (PMAE) are reported. PSD is the Bayesian equivalent of standard error and characterizes the variance of the effect. PMAE is the posterior mean allelic effect; when these values are similar between ancestry cohorts, it suggests that simiar variants are responsible for the effect. When values are in opposite directions, it suggests multiple risk variants, or differing LD structure.
PPH, a posterior probability of heterogeneity; PPA, posterior probability of association; the probability that an SNP is truly associated with a phenotype.
*Proxy for rs2814778.

Conditional analysis

Conditional analysis adjusting for the effect of the most significantly associated SNP at each locus was performed to assess the independence of possible novel variants and to detect the presence of any secondary association signals within known regions. Secondary signals were defined as additional associated variants within 500 kb of previously known loci. Of the primary (known and novel) loci, approximately half contained secondary signals; these include four signals associated with monocytes, three associated with neutrophils and six associated with WBC count. In some instances, the top signal identified by the secondary analysis was stronger than that observed in the primary analysis. This could occur, for example, when the allele frequencies of the initial, index SNP are similar across ancestries, but the conditional signal(s) more accurately tag a functional variant in the respective populations. In order to verify the authenticity of our conditional analysis results, we performed reverse conditioning on our secondary signals and found the signals reported here to remain significant, suggesting an independent effect on WBC subtypes. The top association signals from these conditional analyses are found in Table 4.

Expression quantitative trait loci analysis

All known, conditional and novel loci were assessed as potential expression quantitative trait loci (eQTLs) in leukocyte-derived tissues in order to identify any correlations between association signals and gene expression, as such correlations may account for functional relationships that are not captured by LD. Two known loci associated with monocyte count, and one novel locus newly associated with both WBC and neutrophil count, represented significant ($P < 5E-05$) eQTLs when assessing either the European ancestry sentinel SNPs or their proxies in the YRI (African) and ASN (Asian) populations (1). Three of the four index SNPs at these loci are located within chromosome 6p21.3; each has been associated with a different blood cell trait. The monocyte eQTLs on chromosome six are defined by the transcription factor gene *TCF19*; the neutrophil eQTLs relate primarily to the *HLA* transcripts, but depending on the tissue type are also associated with expression of *ATP6V1G2*, which encodes an enzyme involved in eukaryotic cell compartment acidification. The WBC count eQTLs also relate to various transcripts of this region. Additionally, a previously described locus associated with monocyte count on chromosome 2 is an eQTL for the *ITGA4* transcript at 2q31.3, which encodes the integrin alpha-4 subunit of the very late antigen-4 receptor on monocytes and other mononuclear cells (27,28). In instances where a proxy eQTL SNP has been used to represent a WBC or subtype-associated SNP, variants with greater concordance to the index SNP are viewed with more confidence than those with lower R^2 values. Only two eQTL SNPs were found to be concordant or in very high LD: rs3130320 ($R^2 = 1$) for neutrophils and rs6740847 ($R^2 = 0.983$) for monocytes. We have reported other SNPs as potential eQTLs if their R^2 value with the proxy eQTL SNP is >0.5; these SNPs are of course likely to be near the blood cell SNPs' physical location. All the proxy eQTL variants identified here are located in Supplementary Material, Table S4.

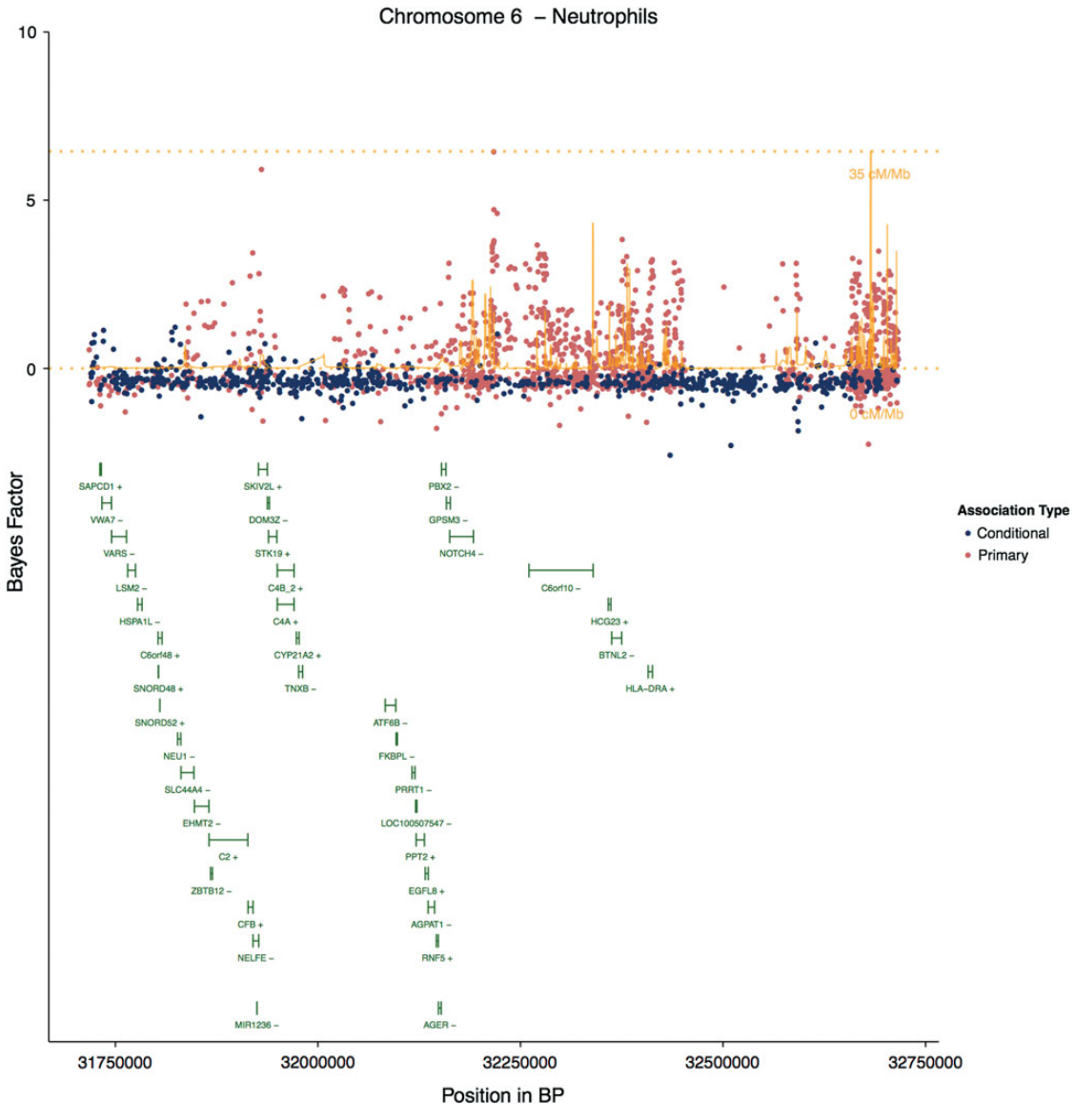

Figure 3. Locus plot for Neutrophil association on chromosome 6. Vertical axis indicates BF, and horizontal axes indicate both chromosomal position and gene location.

DISCUSSION

We applied trans-ethnic MA to summary data from Japanese, African-American and European-Americans populations and identified six new regions that contain biologically plausible genetic loci associated with WBC traits. Many of the novel and secondary association signals we observed involve genomic regions that contain several inflammatory and immune cell-related genes.

Particularly interesting, novel regions include the two loci on chromosome 2 associated with WBC count. The first, identified by rs6734238, falls within an inflammatory gene region of the interleukin-1 cytokine gene family (29). This region has been associated with several inflammation-related biomarkers, including C-reactive protein and fibrinogen. rs6734238 is located downstream of *IL1F10* and upstream of *IL1RN*. *IL1RN* encodes IL-1 receptor antagonist (IL-1RA), which regulates a variety of interleukin-1-related immune and inflammatory responses, including inhibition of interleukin 1, alpha (IL1A) and interleukin 1, beta (IL1B). *IL1F10* encodes IL-38, which regulates Th17 immune responses and stimulates IL-6 cytokine production from dendritic cells *in vitro* (30). The second chromosome 2 region, identified by rs10932765, is near *APRC2* and *CXCR1* (also known as *IL8RA*). *CXCR1* is of particular interest as this is a chemokine receptor involved in leukocyte chemotaxis and trafficking (31).

The novel chromosome 5q13 region associated with monocyte count lies within a family of protease-activated receptor genes, *F2RL2-F2RL1-F2R*. The *F2RL1* gene (protease-activated receptor-2 or PAR-2) has previously been related to some inflammatory and autoimmune diseases, and is a known receptor on monocytes (32). PAR-2 is a G protein-coupled receptor on monocyte/macrophages and other cell types that appear to have a direct role in the regulation of innate immune

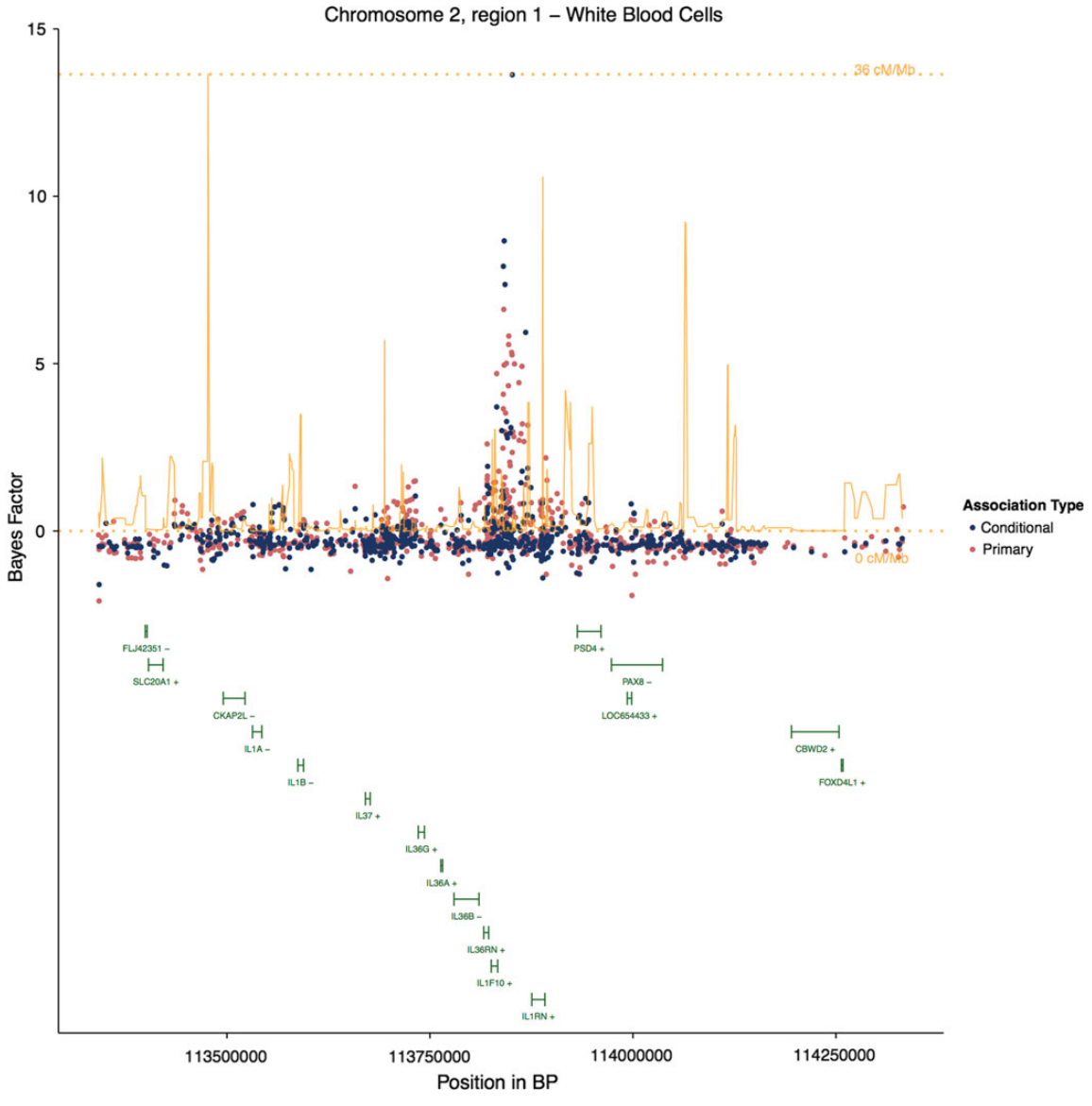

Figure 4. Locus plot for White Blood Cell count association on chromosome 2 (first region). Vertical axis indicates BF, and horizontal axes indicate both chromosomal position and gene location.

function. Specifically, PAR-2 can be activated by a number of endogenous inflammation-associated proteinases (e.g. mast cell tryptase, trypsin and neutrophil proteinase 3) or exogenous pathogen-derived proteinases (33). Notably, another member of the PAR gene family located on 5q13, *F2R*, encodes PAR-1, the platelet thrombin receptor. Common variants of *F2R* were recently associated with circulating platelet count in a European GWAS (23).

A single region of chromosome 6, associated with neutrophil and WBC count, is located near the *HBS1L* and *MYB* genes, which are known to be associated with fetal hemoglobin levels and monocyte counts. This locus has also been reported to be associated with red cell and platelet traits, but not previously with white cell traits. Although there is a single signal for neutrophils at this locus, the signal for WBC count appears bimodal. One of these two regions is captured by a single SNP with 99% confidence, while the second requires a regional span of nearly 1 Mb to reach the same level of confidence. This locus was also significant in our eQTL analysis across all ancestry types for both neutrophil and WBC count, in lymphoblastoid cell lines (LCLs) and whole blood.

The region of chromosome 8 newly associated with WBC count lies near a gene of unknown function, *GSDMC*, which encodes gasdermin C. Other genes belonging to the gasdermin family have been associated with immune-mediated phenotypes such as asthma and alopecia (34,35), suggesting a role for this gene family in inflammatory disorders (36).

A signal was observed on chromosome 16 for WBC count and neutrophils; however, the top associated variant is located within the intronic region of *HYDIN*, an mRNA transcript sequence involved in cilia motility. As previously reported, it is a likely homolog to the *DARC* region of chromosome 1 and represents a spurious signal (12).

The replicated region on chromosome 4, associated with both WBC and neutrophil counts, is located near a chemokine family

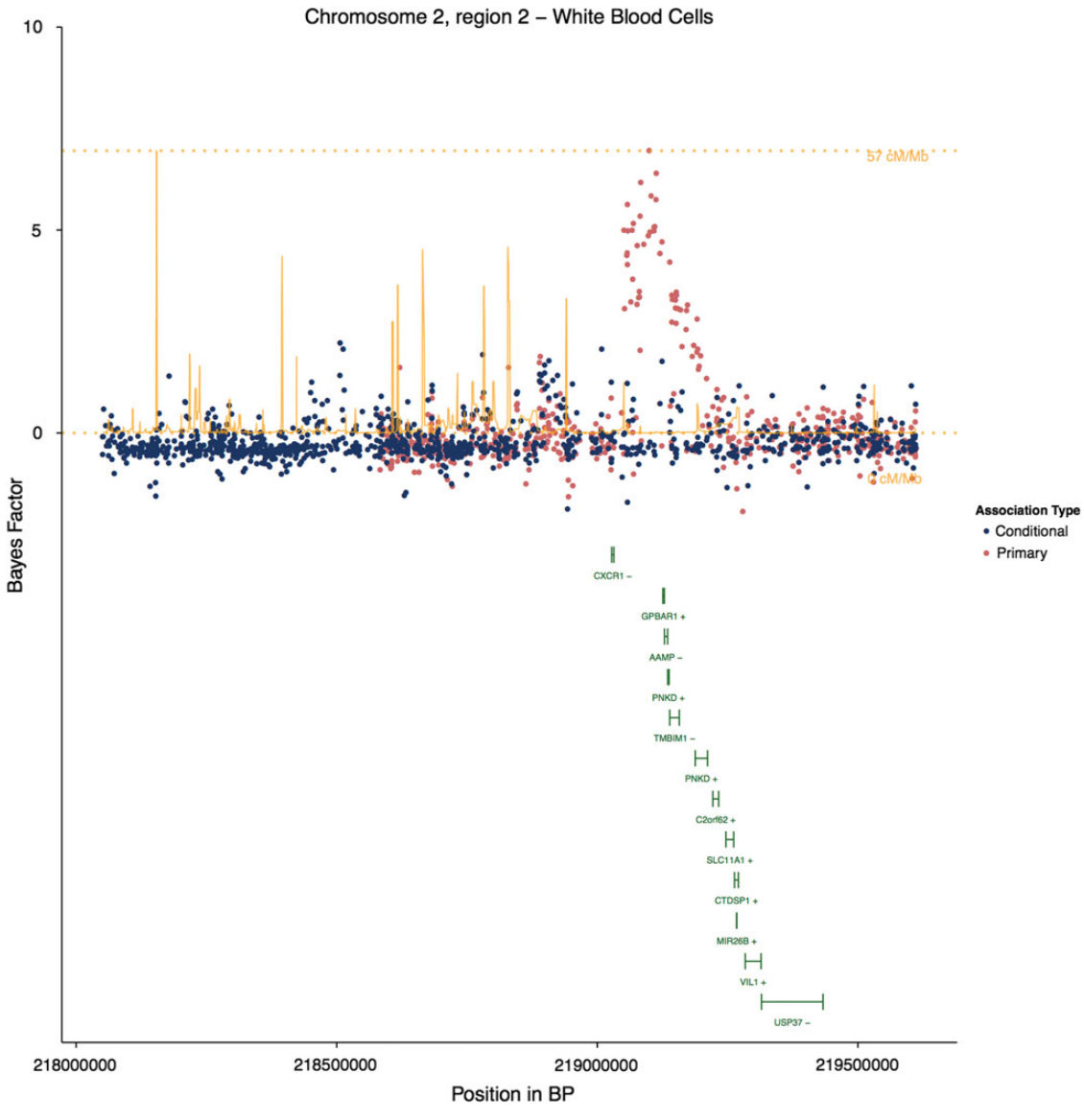

Figure 5. Locus plot for White Blood Cell count association on chromosome 2 (second region). Vertical axis indicates BF, and horizontal axes indicate both chromosomal position and gene location.

gene cluster, *CXCL5-CXCL3-CXCL2*. *CXCL2* interacts with another chemokine receptor, *CXCR2*, to control migration of leukocytes from the bone marrow (37). Taken together with the *DARC* locus, these findings extend the importance of common genetic variants of chemokine ligands and receptors in the regulation of WBC counts.

Using the proper reference panel is critical to the dependability and accuracy of this analysis. For this reason, ancestry-matched subsets of the 1000 Genomes were used as the reference panel from which LD was calculated, which are presumed to be drawn from the same general populations as those used here. However, the relatively small sample sizes available through 1000 Genomes increase the possibility of error in LD estimations (38). While the localization and resolution of functional variants may improve with the additional genomic variation measured in newer reference panels, we show that even without these newer panels, the associations identified using MANTRA provide plausible candidates for functionality.

In conclusion, trans-ethnic meta-analyses allow for an examination of disease traits within a large population of individuals and provide the opportunity to localize previously known regions and detect novel ones, while considering the heterogeneity of allelic effects that may exist between contentially distinct populations. Additionally, our results illustrate the utility of trans-ethnic fine mapping for narrowing regions of association. Well-established loci replicated in the present study show credible intervals that flank the known index variant. For example, the previously known monocyte associations on chromosomes 8 and 9, shown in Supplementary Material, Table S3, have 99% credible intervals of only a few thousand base-pairs, located in close proximity to the originally identified variants. Our analysis of the *DARC* region is complicated by the causal

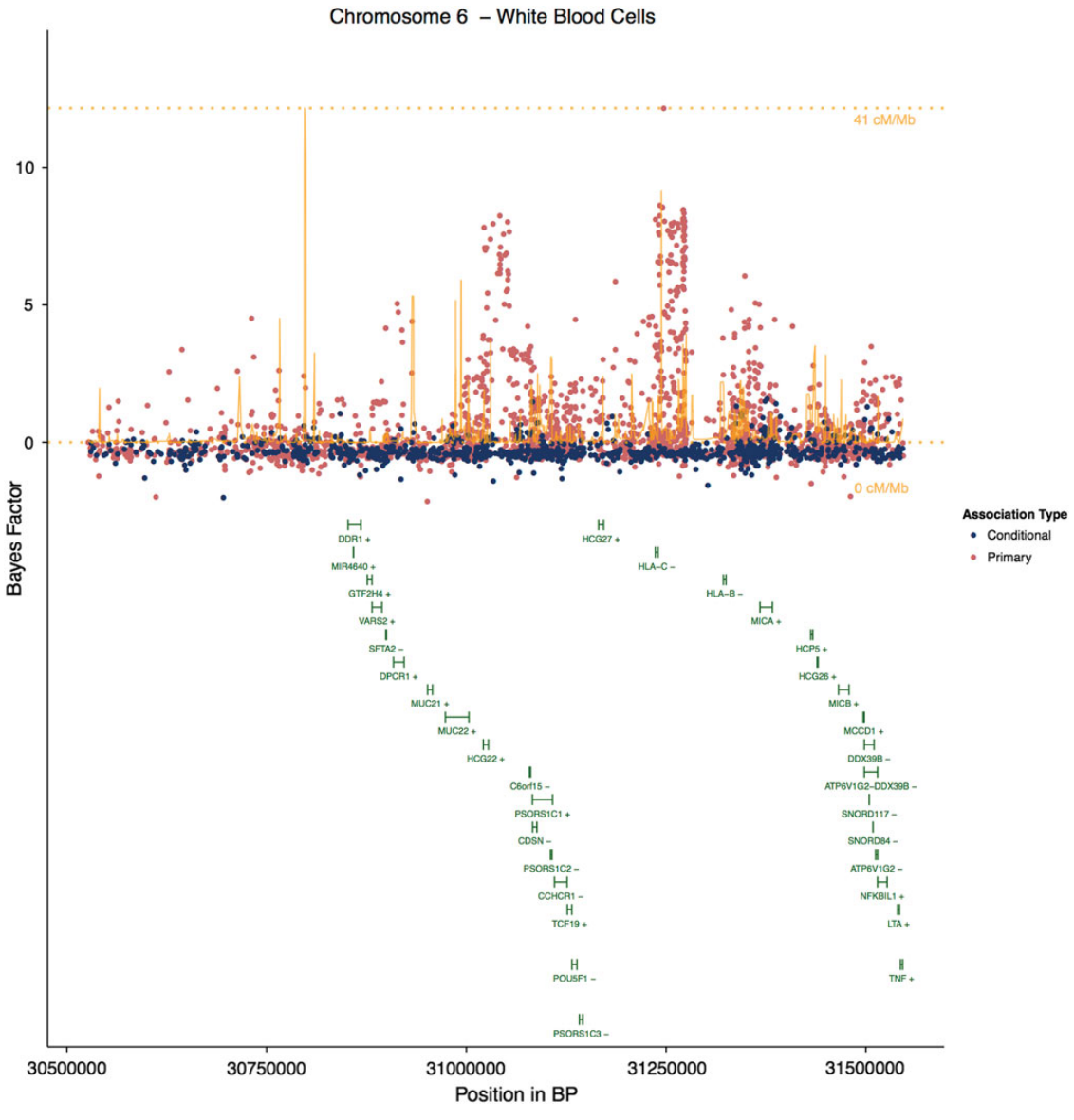

Figure 6. Locus plot for White Blood Cell count association on chromosome 6. Vertical axis indicates BF, and horizontal axes indicate both chromosomal position and gene location.

SNP being monomorphic in two of the three populations employed. When the credible interval analysis is applied to this region, the proxy variant (rs2518564) tagging the known functional variant (rs2814778) (5,8) is identified as accounting for the entire signal. However, prior association studies and evidence of biological function allow confident identification of rs2814778 as the functional variant. By calculating credible intervals across test statistics from analyses of combined ancestries, we were able to narrow expansive loci to smaller regions. Further work is necessary to identify what functional variants may lie within these regions.

The increasing availability of GWAS summary data for many phenotypic traits of interest, from many ethnically diverse populations, suggests that the trans-ethnic GWAS MA approach can yield additional association signals, thereby explaining some of the missing heritability and genetic architecture for other complex traits. In addition, this work is relevant for future targeted sequencing follow-up studies, as we have narrowed the scope of follow-up sequencing efforts for functional variants. By increasing the mapping resolution of the causal variants within these loci, we hope that these results guide next-generation targeted deep sequencing studies, which may disentangle the heterogeneity of effect across ethnicities (39). Future work will discern which functional variants are the same across ethnicities and which tag nearby regions, through LD, that harbor the true functional variant or variants.

MATERIALS AND METHODS

The trans-ethnic GWA strategy was applied to three consortia containing WBC phenotypes. These include a Japanese population represented by RIKEN (Rikagaku Kenkyusho, Institute of

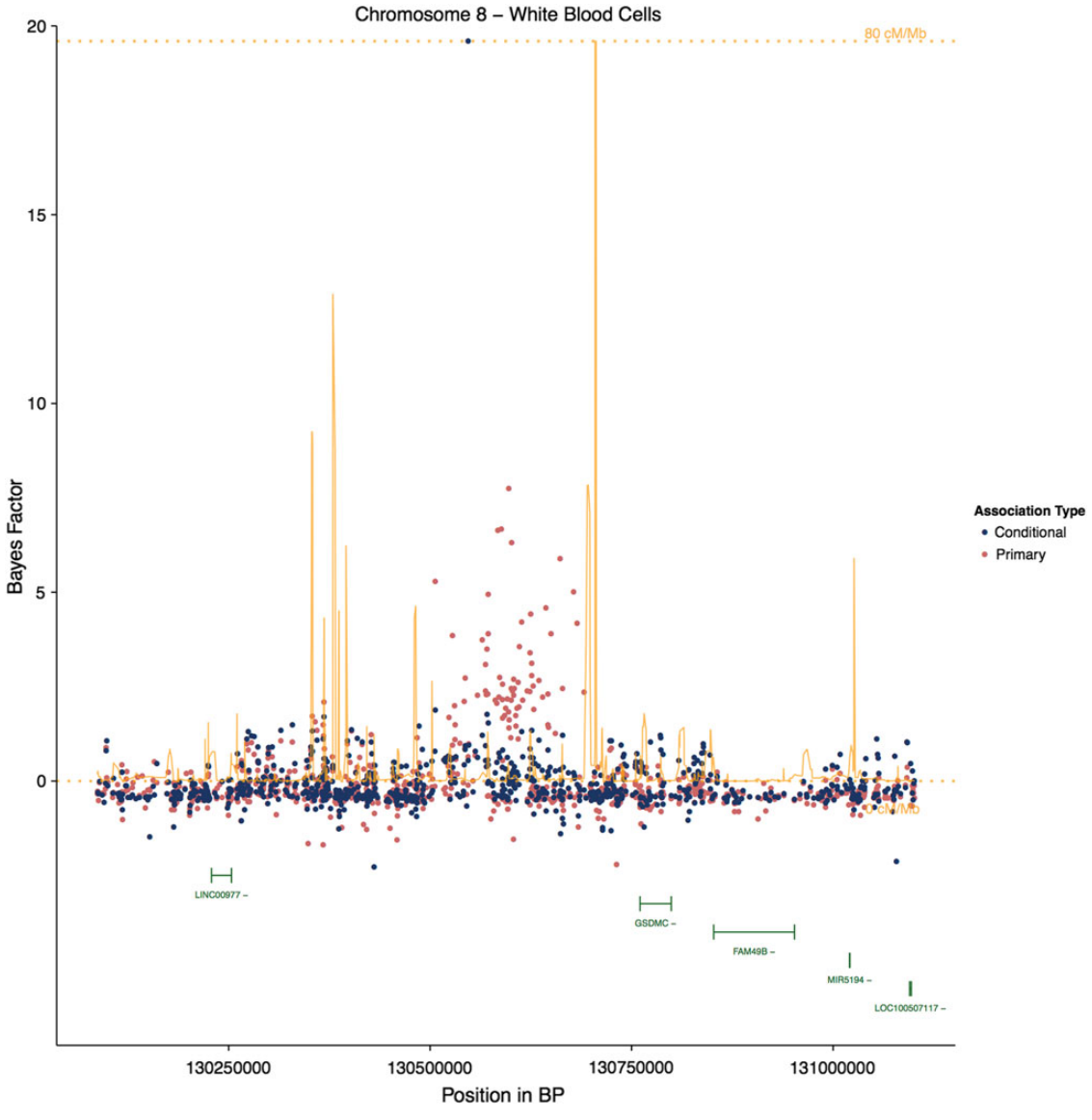

Figure 7. Locus plot for White Blood Cell count association on chromosome 8. Vertical axis indicates BF, and horizontal axes indicate both chromosomal position and gene location.

Physical and Chemical Research, Japan), a European ancestry population represented by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium and an African-American population represented by the Continental Origins and Genetic Epidemiology Network (COGENT) Consortium (9,11,12). All three consortia contain measurements for total WBC count, neutrophil count and monocyte count, using the same scale of transformation and similar analytic paradigms. Clinical information of the subjects includes age, gender and smoking history, and was collected by self-report. Subject BMI was also collected as a measure. The laboratory data include total WBC count and subtypes, as determined using automated hematology cell counters according to the standardized protocol. The WBC phenotypes were natural log-transformed prior to analysis to provide a normal distribution. Samples >2 SD outside of the ethnicity specific mean for the given phenotype were excluded. This was done to ensure normality of the included samples, and to exclude any subclinical inflammation such as the common cold. If a sample was identified as an outlier for one or more subtype, it was excluded entirely from the study.

The RIKEN study comprises over 17 000 individuals from The BioBank Japan Project, which is made of up over 300 000 subjects (http://biobankjp.org) (11). Samples determined to be of non-Japanese origin by either self-report or by principal components analyses (PCAs) in GWAS were excluded from further analyses. For the GWAS, 592 232 SNPs were genotyped using Illumina HumanHap610-Quad Genotyping BeadChip. Subjects with call rates <0.98 were removed, as were SNPs with call rates <0.99. First- and second-degree relatives were excluded based on identity-by-descent analyses, as were SNPs with minor allele frequency (MAF) <0.01 or with Hardy–Weinberg equilibrium (HWE) P-values $<1.00\text{E}-7$. After quality control, genotypes were imputed using MACH 1.0 in a two-step procedure,

Table 4. Loci identified by conditional analysis

Subtype	Primary top hits[a]			Secondary top hits								
	SNP	Chr	Position	SNP	Position	Effect allele	Other allele	*N*—studies	$\log_{10}$ BF	PPH	Sample size	Effect direction
MONO	rs1449263	2	182 319 301	rs711801	182 334 873	C	T	3	10.46982	0.579	33 711	++−
MONO	rs3095254	6	31 221 668	rs2517774	29 893 982	C	T	2	72.9753	1	26 360	?−−
MONO	rs2163952	8	130 610 389	rs1457475	134 988 329	A	G	3	6.48727	1	28 317	−−+
MONO	rs12350763	9	113 923 723	rs4401938	121 179 496	C	T	3	68.41786	0.99	33 729	+++
NEU	rs7667376	4	74 967 890	rs1440404	74 944 449	C	G	3	54.50984	1	33 713	++−
NEU	rs445	7	92 408 370	rs3731326	92 327 026	A	G	3	6.86032	1	33 753	+−−
NEU	rs4794822	17	38 156 712	rs4794321	46 028 844	C	T	3	7.89538	1	33 753	−−+
WBC	rs6734238	2	113 841 030	rs11899198	113 840 539	G	T	3	7.90504	1	52 694	−++
WBC	rs1371799	4	74 977 837	rs1440404	74 944 449	C	G	3	53.18811	1	52 697	+−−
WBC	rs9402686	6	135 427 817	rs1890428	140 253 023	C	T	2	22.53076	1	22 472	+?−
WBC	rs445	7	92 408 370	rs42626	89 961 237	C	T	3	11.88853	1	52 740	−++
WBC	rs2163950	8	130 597 585	rs10505542	130 547 253	C	T	3	19.59631	1	52 686	−−+
WBC	rs2241245	17	38 151 014	rs8070454	38 160 754	C	T	2	22.03367	1	33 231	+?−

Top hits from the GCTA conditional analyses are reported.
PPH, posterior probability of association; or, the probability that an SNP is truly associated with a phenotype. Effect direction order of studies: COGENT, CHARGE, RIKEN.
[a]Used as covariates in conditional analysis.

described in detail elsewhere (40,41). HapMap Phase II Japanese individuals from Tokyo (JPT) and Han Chinese individiuals from Beijing (CHB) individuals were adopted as references. SNPs with imputation qualities <0.30 were excluded prior to analyses, and genomic control was applied to the cohort-level data. All participants provided written informed consent as approved by the ethical committees of the Center for Genomic Medicine, RIKEN and the Institute of Medical Science, the University of Tokyo (11).

The CHARGE consortium dataset is comprised of over 19 000 individuals from seven discovery cohorts, including: the Rotterdam Study (RS), Framingham Heart Study (FHS), the NHLBI's Atherosclerosis Risk in Communities (ARIC) Study, the Age, Gene/Environment Susceptibility—Reykjavik Study (AGES), Health Aging and Body Composition study (HABC), the Baltimore Longitudinal Study of Aging (BLSA) and the Invecchaire in Chianti Study (inChianti) (9). Each of these studies, with the exception of the Framingham Heart Study, is comprised of unrelated individuals of confirmed European ancestry, based on PCAs. Prior to MA, SNPs with MAF <0.01, missingness $>5\%$ or HWE $<1.00\text{E}-7$ were excluded. Individuals with call rates $<95\%$ were also excluded. After quality control, genotypes were imputed using the CEU reference panel of the HapMap Phase II haplotype data. The CHARGE consortium is comprised of MA data resulting from the summary statistics of these individual studies. Prior to the meta- analyses, study results were adjusted for genomic inflation factors, and SNPs with imputation quality <0.30 were excluded. Meta-analyses were performed using a fixed-effects model in METAL (42).

The COGENT consortium is comprised of over 16 000 self-identified African-Americans from seven discovery cohorts, including: Atherosclerosis Risk in Communities (ARIC), Coronary Artery Risk Development in Young Adults (CARDIA), Johns Hopkins Genetic Study of Atherosclerosis Risk (GeneSTAR), HealthyAging in Neighborhoods of Diversity across the Life Span (HANDLS), Health, Aging, and Body Composition (Health ABC), Jackson Heart Study (JHS) and the Women's Health Initiative (WHI) (12). SNPs were excluded from cohort-level GWAS if MAF $<1\%$ or missingness $>5\%$. Monomorphic SNPs and ambiguously mapped SNPs were also removed. Individual samples exhibiting gender mismatch or genotype missingness $>10\%$ were excluded. After quality control, genotypes were imputed with HapMap Phase II, using a 1 : 1 mixture of the CEU and YRI reference populations. Prior to MA, SNPs with imputation quality <0.30 were excluded. Study-specific GWA results were corrected for genomic inflation factors, and MA was performed using a fixed-effects model in METAL (42).

Statistical analysis

Summary statistics for the RIKEN GWAS and the CHARGE and COGENT meta-analyses were collected and stratified by ethnicity and WBC subtype availability. Data were input into the trans-ethnic MA software package, MANTRA (15), which makes use of a prior model of relatedness between studies corresponding to Fst, or mean effect allele frequency differences between populations. Relatedness is determined by differences in allele frequency between studies. MANTRA estimates the BF in favor of association for each SNP using a Markov chain Monte Carlo (MCMC) algorithm. Results are reported as $\log_{10}$ (BF), and associations of 6 or greater have the highest posterior odds of being truly present (43,44). Posterior probability of heterogeneity is also reported to examine levels of variation in allelic effects across the populations used in the analysis. Combining results across studies using a Bayesian approach is advantageous, as the evidence produced by this study is directly comparable to future studies performed in the same way. Simulations of distinct ancestry populations show that when MANTRA is compared with random-effects and fixed-effects meta-analyses, MANTRA shows increased performance and produces the highest-powered results in the detection of novel associations (15,22,45–47).

Since the initial imputation of the datasets used here, a number of more comprehensive reference panels have been released including the latest HapMap release and the samples available

through the 1000 Genomes Project. As these reference panels contain more individuals and greater genome coverage, more genotypes are predicted with greater confidence than when using prior HapMap releases. Imputing raw data to the latest release of 1000 Genomes would be ideal; however, due to the data-sharing requirements of the cohorts included in this analysis, only summary statistics were available for these datasets. However, MANTRA is still expected to outperform a traditional MA in this case, as MANTRA accounts for heterogeneity while making no assumptions about differences or similarities in allelic effect.

To quantify uncertainty surrounding the top hits from the trans-ethnic MANTRA analysis, we calculated 95 and 99% credible regions (48). We estimated credible sets of SNPs by first defining a 1 Mb genomic region surrounding lead SNPs ($\pm$ 500 kb), then ranking the regional SNPs within this region according to their BF and then combining the cumulative posterior probabilities of these ranked SNPs until 95 and 99% confidence was reached.

In addition, the top hits from MANTRA were input into a conditional analysis in order to identify additional association signals at nearby susceptibility loci and to determine independence of these secondary signals from the index SNP association. As complex diseases are assumed to be influenced by two or more genes acting in concert, it is possible that prior GWAS aimed at identifying single loci have not detected secondary signals. Thus, when large sample sizes are available, conditional approaches can be useful in detecting secondary association signals with loci that initially appear to contribute a negligible risk to disease susceptibility. In addition, as evidence of association is predicated on a given conditioned SNP, it is possible that, in some instances, the secondary associations are stronger than what was observed in the primary single-ethnicity GWAS analysis. This is possible when the allele frequencies of a given SNP are similar across ancestries, as conditional hits may be closer to a functional variant than the original, single-population hit.

We used the software program Genome-wide Complex Trait Analysis (GCTA) v1.13 to perform conditional association analysis for each ancestry-specific set of summary results (49). As individual genotype data were unavailable, this was performed separately for each cohort using summary statistics and incorporating LD information from ancestry-matched reference samples containing individual-level genotype data. When original genotype data are not available, it is essential that the reference samples be from the same population as the original data, so that the LD structure estimated from the reference population is not biased. It is also critical that the reference sample is not affected by cryptic relatedness or population stratification. This is particularly relevant to admixed populations, such as COGENT, which is comprised of African-American individuals. In order to avoid confounding the genetic relationship matrix (GRM) produced by GCTA, eigenvectors are included in the model as covariates, to capture and account for any variance that is present due to population structure (49). In this analysis, we used ancestry-matched subsets from the 1000 Genomes Project to estimate LD structure within our samples (17). Using these samples as LD proxies, the GCTA association analyses were conditioned on the top hits from MANTRA, specific to each locus of interest. Independently associated SNPs were selected using a stepwise model selection procedure. Analyses were performed separately for each ancestry cohort. The results from each cohort were then meta-analyzed across ethnicities.

In addition to the trans-ethnic association analysis, a secondary analysis exploring eQTLs was performed using SNAP (50). SNAP is a web server that identifies and annotates nearby proxy SNPs in LD (according to HapMap) to those queried. Using ancestry-specific tissues, SNAP was identified alias SNPs for significant index SNPs, and proxy SNPs in high linkage disequilibrium ($R^2 > 0.5$). Sentinel, alias and proxy SNPs were queried within a collected database of expression SNP (eSNP) results, drawn from the following leukocyte-derived tissues: fresh lymphocytes (51), fresh leukocytes (52), leukocyte samples in individuals with celiac disease (53), whole blood samples (54–56), LCLs derived from asthmatic children (57,58), HapMap LCL from three populations (59), a separate study on HapMap CEU LCL (60) and additional LCL population samples [(61–63); Mangravite *et al*., unpublished], CD19+ B cells (64), primary PHA (phytohaemagglutinin)-stimulated T cells (61), CD4+ T cells (65), peripheral blood monocytes (27,64,66), CD11+ dendritic cells before and after *Mycobacterium tuberculosis* infection (67) and micro-RNA QTLs queried for LCL (68). The collected eSNP results met criteria for statistical thresholds of association with gene transcript levels, as described in the original cited papers. In cases where an SNP was associated with a transcript, we further examined the strongest eSNP for the transcript within that dataset and the LD between the strongest eSNP and blood count-selected eSNPs. This was done to assess the concordance of the blood count and expression signals.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov).

Conflict of Interest statement. None declared.

FUNDING

Y.O. is supported by a grant from the Japan Society of the Promotion of Science (JSPS). The development of the software package MANTRA was supported by Wellcome Trust (grant nos WT098017, WT090532 and WT064890). S.K.G. was supported in part by P30HL101290. Funding for CHARGE was made possible by NIA/NIH contract AG000932-2 (2009) Characterization of Normal Genomic Variability. The Age, Gene/Environment Susceptibility Reykjavik Study is funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association) and the Althingi (the Icelandic Parliament). The Atherosclerosis Risk in the Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C and HHSN268

201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract (U01HG 004402) and National Institutes of Health contract (HHSN 268200625226C). The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by grant number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The National Heart, Lung, and Blood Institute's Framingham Heart Study is a joint project of the National Institutes of Health and Boston University School of Medicine and was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (contract No. N01-HC-25195) and its contract with Affymetrix for genotyping services (contract No. N02-HL-6-4278). Analyses reflect the efforts and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at the Boston University School of Medicine and Boston Medical Center. The Health ABC Study was supported in part by the Intramural Research Program of the NIH, National Institute on Aging, NIA contracts N01AG62101, N01AG62103 and N01AG 62106. The GWAS was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University (contract number HHSN268200782096C). The InChianti Study was supported as a 'targeted project' (ICS 110.1RS97.71) by the Italian Ministry of Health, by the US National Institute on Aging (contracts N01-AG-916413, N01-AG-821336, 263 MD 9164 13 and 263 MD 821336) and in part by the Intramural Research Program, National Institute on Aging, National Institutes of Health, USA. The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organization of Scientific Research NWO Investments (no. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), EUROSPAN (European Special Populations Research Network; LSHG-CT-2006-01947), the Netherlands Organization for Scientific Research (Pionier, 047.016.009, 047.017. 043;050-060-810), Erasmus Medical Center and the Centre for Medical Systems Biology (CMSB I and II and Grand; National Genomics Initiative) of the Netherlands Genomics Initiative (NGI); The Rotterdam Study is further funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII) and the Municipality of Rotterdam. Funding for COGENT was obtained through the Broad Institute (N01-HC-65226) to create this genotype/phenotype database for wide dissemination to the biomedical research community. The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268 201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN26820 1100011C and HHSN268201100012C), R01HL087641, R01HL 59367 and R01HL086694; National Human Genome Research Institute (contract U01HG004402) and National Institutes of Health (contract HHSN268200625226C). The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by grant number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. ARIC is also supported by University of North Carolina at Chapel Hill (funded by N01-HC-55015), Baylor Medical College (N01-HC-55016), University of Mississippi Medical Center (N01-HC-55021), University of Minnesota (N01-HC-55019), Johns Hopkins University (N01-HC-55020), University of Texas Houston (N01-HC-55017) and University of North Carolina (N01-HC-55018). Coronary Artery Risk in Young Adults (CARDIA): University of Alabama at Birmingham (N01-HC-48047), University of Minnesota (N01-HC-48048), Northwestern University (N01-HC-48049), Kaiser Foundation Research Institute (N01-HC-48050), University of Alabama at Birmingham (N01-HC-95095), Tufts-New England Medical Center (N01-HC-45204), Wake Forest University (N01-HC-45205), Harbor-UCLA Research and Education Institute (N01-HC-05187), University of California, Irvine (N01-HC-45134 and N01-HC-95100). Jackson Heart Study (JHS): Jackson State University (N01-HC-95170), University of Mississippi (N01-HC-95171), Tougaloo College (N01-HC-95172). Healthy Aging in Neighborhoods of Diversity across the Life Span Study (HANDLS): This research was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (intramural project # Z01-AG000513 and human subjects protocol # 2009-149). Health ABC: This research was supported by NIA contracts N01AG62101, N01AG62103 and N01AG62106. The GWAS was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University (contract number HHSN268200782096C). This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. GeneSTAR Acknowledgement: This research was supported by the National Heart, Lung, and Blood Institute (NHLBI) through the PROGENI (U01 HL72518) and STAMPEED (R01 HL087698-01) consortia. Additional support was provided by grants from the NIH/National Institute of Nursing Research (R01 NR08153) and the NIH/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center. WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services, through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32 and 44221. The Cardiovascular Health Study (CHS) is supported from contracts N01-HC-35129, N01-HC-45133, N01-HC-75150, N01-HC-85079 through N01-HC-85086, N01 HC-15103, N01 HC-55222 and U01 HL080295 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support for this

work was provided by NIH R01 HL71862-06 and ARRA N000949304 (to A.P.R.). Some of the results of this paper were obtained by using the program package S.A.G.E., which is supported by a US Public Health Service Resource Grant (RR03655) from the National Center for Research Resources. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Funding for RIKEN and the BioBank Japan Project was supported by Ministry of Education, Culture, Sports, Science and Technology, Japan. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

1. Ruggiero, C., Metter, E.J., Cherubini, A., Maggio, M., Sen, R., Najjar, S.S., Windham, G.B., Ble, A., Senin, U. and Ferrucci, L. (2007) White blood cell count and mortality in the Baltimore Longitudinal Study of Aging. *J. Am. Coll. Cardiol.*, **49**, 1841–1850.
2. Danesh, J., Collins, R., Appleby, P. and Peto, R. (1998) Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. *JAMA*, **279**, 1477–1482.
3. Shankar, A., Wang, J.J., Rochtchina, E., Yu, M.C., Kefford, R. and Mitchell, P. (2006) Association between circulating white blood cell count and cancer mortality: a population-based cohort study. *Arch. Intern. Med.*, **166**, 188–194.
4. Nieto, F.J., Szklo, M., Folsom, A.R., Rock, R. and Mercuri, M. (1992) Leukocyte count correlates in middle-aged adults: the Atherosclerosis Risk in Communities (ARIC) Study. *Am. J. Epidemiol.*, **136**, 525–537.
5. 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.
6. Pilia, G., Chen, W.-M., Scuteri, A., Orrú, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P. *et al.* (2006) Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.*, **2**, e132.
7. Haddy, T.B., Rana, S.R. and Castro, O. (1999) Benign ethnic neutropenia: what is a normal absolute neutrophil count? *J. Lab. Clin. Med.*, **133**, 15–22.
8. Reich, D., Nalls, M.A., Kao, W.H.L., Akylbekova, E.L., Tandon, A., Patterson, N., Mullikin, J., Hsueh, W.-C., Cheng, C.-Y., Coresh, J. *et al.* (2009) Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet.*, **5**, e1000360.
9. Nalls, M.A., Couper, D.J., Tanaka, T., van Rooij, F.J.A., 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.
10. Okada, Y. and Kamatani, Y. (2012) Common genetic factors for hematological traits in humans. *J. Hum. Genet.*, **57**, 161–169.
11. Okada, Y., Hirota, T., Kamatani, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Higasa, K., Yamaguchi-Kabata, Y., Hosono, N., Nalls, M.A. *et al.* (2011) Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS Genet.*, **7**, e1002067.
12. 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.
13. 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.
14. 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.
15. Morris, A.P. (2011) Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.*, **35**, 809–822.
16. International HapMap Consortium (2003) The International HapMap Project. *Nature*, **426**, 789–796.
17. 1000 Genomes Project ConsortiumAbecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
18. Mägi, R. and Morris, A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, **11**, 288.
19. Han, B. and Eskin, E. (2011) Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 586–598.
20. Tian, C., Gregersen, P.K. and Seldin, M.F. (2008) Accounting for ancestry: population substructure and genome-wide association studies. *Hum. Mol. Genet.*, **17**, R143–e150.
21. Franke, A., Balschun, T., Sina, C., Ellinghaus, D., Häsler, R., Mayr, G., Albrecht, M., Wittig, M., Buchert, E., Nikolaus, S. *et al.* (2010) Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL). *Nat. Genet.*, **42**, 292–294.
22. Dastani, Z., Hivert, M.-F., Timpson, N., Perry, J.R.B., Yuan, X., Scott, R.A., Henneman, P., Heid, I.M., Kizer, J.R., Lyytikäinen, L.-P. *et al.* (2012) Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.*, **8**, e1002607.
23. Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A.H., Labrune, Y. *et al.* (2011) New gene functions in megakaryopoiesis and platelet formation. *Nature*, **480**, 201–208.
24. Johansson, U., Lawson, C., Dabare, M., Syndercombe-Court, D., Newland, A.C., Howells, G.L. and Macey, M.G. (2005) Human peripheral blood monocytes express protease receptor-2 and respond to receptor activation by production of IL-6, IL-8, and IL-1{beta}. *J. Leukoc. Biol.*, **78**, 967–975.
25. Stephens, M. and Balding, D.J. (2009) Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.*, **10**, 681–690.
26. Hamblin, M.T. and Di Rienzo, A. (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.*, **66**, 1669–1679.
27. Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., Maouche, S., Germain, M., Lackner, K., Rossmann, H. *et al.* (2010) Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE*, **5**, e10693.
28. Maugeri, N., Powell, J.E., 't Hoen, P.A.C., de Geus, E.J.C., Willemsen, G., Kattenberg, M., Henders, A.K., Wallace, L., Penninx, B., Hottenga, J.-J. *et al.* (2011) LPAR1 and ITGA4 regulate peripheral blood monocyte counts. *Hum. Mutat.*, **32**, 873–876.
29. Van de Veerdonk, F.L. and Netea, M.G. (2013) New insights in the immunobiology of IL-1 family members. *Front. Immunol.*, **4**, 167.
30. Van de Veerdonk, F.L., Stoeckman, A.K., Wu, G., Boeckermann, A.N., Azam, T., Netea, M.G., Joosten, L.A.B., van der Meer, J.W.M., Hao, R., Kalabokis, V. *et al.* (2012) IL-38 binds to the IL-36 receptor and has biological effects on immune cells similar to IL-36 receptor antagonist. *Proc. Natl. Acad. Sci. USA*, **109**, 3001–3005.
31. Raghuwanshi, S.K., Su, Y., Singh, V., Haynes, K., Richmond, A. and Richardson, R.M. (2012) The chemokine receptors CXCR1 and CXCR2 couple to distinct G protein-coupled receptor kinases to mediate and regulate leukocyte functions. *J. Immunol.*, **189**, 2824–2832.
32. Rothmeier, A.S. and Ruf, W. (2012) Protease-activated receptor 2 signaling in inflammation. *Semin. Immunopathol.*, **34**, 133–149.
33. Shpacovitch, V., Feld, M., Bunnett, N.W. and Steinhoff, M. (2007) Protease-activated receptors: novel PARtners in innate immunity. *Trends Immunol.*, **28**, 541–550.
34. Li, X., Ampleford, E.J., Howard, T.D., Moore, W.C., Torgerson, D.G., Li, H., Busse, W.W., Castro, M., Erzurum, S.C., Israel, E. *et al.* (2012) Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases. *J. Allergy Clin. Immunol.*, **130**, 861–868.e7.
35. Zhou, Y., Jiang, X., Gu, P., Chen, W., Zeng, X. and Gao, X. (2012) Gsdma3 mutation causes bulge stem cell depletion and alopecia mediated by skin inflammation. *Am. J. Pathol.*, **180**, 763–774.
36. Saeki, N., Usui, T., Aoyagi, K., Kim, D.H., Sato, M., Mabuchi, T., Yanagihara, K., Ogawa, K., Sakamoto, H., Yoshida, T. *et al.* (2009) Distinctive expression and function of four GSDM family genes

(GSDMA-D) in normal and malignant upper gastrointestinal epithelium. *Genes Chromosomes Cancer*, **48**, 261–271.
37. Day, R.B. and Link, D.C. (2012) Regulation of neutrophil trafficking from the bone marrow. *Cell. Mol. Life Sci.*, **69**, 1415–1423.
38. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J. *et al.* (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.*, **44**, 369–375.
39. Goldstein, D.B. (2011) The importance of synthetic associations will only be resolved empirically. *PLoS Biol.*, **9**, e1001008.
40. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. and Abecasis, G.R. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, **44**, 955–959.
41. Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
42. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
43. Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.*, **11**, 241–247.
44. Sawcer, S. (2010) Bayes factors in complex genetics. *Eur. J. Hum. Genet.*, **18**, 746–750.
45. Wang, X., Chua, H.-X., Chen, P., Ong, R.T.-H., Sim, X., Zhang, W., Takeuchi, F., Liu, X., Khor, C.-C., Tay, W.-T. *et al.* (2013) Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum. Mol. Genet.*, **22**, 2303–2311.
46. Urbanek, M., Hayes, M.G., Armstrong, L.L., Morrison, J., Lowe, L.P., Badon, S.E., Scheftner, D., Pluzhnikov, A., Levine, D., Laurie, C.C. *et al.* (2013) The chromosome 3q25 genomic region is associated with measures of adiposity in newborns in a multi-ethnic genome-wide association study. *Hum. Mol. Genet.*, **22**, 3583–3596.
47. Franceschini, N., Fox, E., Zhang, Z., Edwards, T.L., Nalls, M.A., Sung, Y.J., Tayo, B.O., Sun, Y.V., Gottesman, O., Adeyemo, A. *et al.* (2013) Genome-wide association analysis of blood-pressure traits in African-Ancestry individuals reveals common associated genes in African and non-African populations. *Am. J. Hum. Genet.* 10.1016/j.ajhg.2013.07.010.
48. Franceschini, N., van Rooij, F.J.A., Prins, B.P., Feitosa, M.F., Karakas, M., Eckfeldt, J.H., Folsom, A.R., Kopp, J., Vaez, A., Andrews, J.S. *et al.* (2012) Discovery and fine mapping of serum protein loci through transethnic meta-analysis. *Am. J. Hum. Genet.*, **91**, 744–753.
49. 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.
50. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I.W. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, **24**, 2938–2939.
51. Göring, H.H.H., Curran, J.E., Johnson, M.P., Dyer, T.D., Charlesworth, J., Cole, S.A., Jowett, J.B.M., Abraham, L.J., Rainwater, D.L., Comuzzie, A.G. *et al.* (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.*, **39**, 1208–1216.
52. Idaghdour, Y., Czika, W., Shianna, K.V., Lee, S.H., Visscher, P.M., Martin, H.C., Miclaus, K., Jadallah, S.J., Goldstein, D.B., Wolfinger, R.D. *et al.* (2010) Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nat. Genet.*, **42**, 62–67.
53. Heap, G.A., Trynka, G., Jansen, R.C., Bruinenberg, M., Swertz, M.A., Dinesen, L.C., Hunt, K.A., Wijmenga, C., Vanheel, D.A. and Franke, L. (2009) Complex nature of SNP genotype effects on gene expression in primary human leucocytes. *BMC Med. Genomics*, **2**, 1.
54. Fehrmann, R.S.N., Jansen, R.C., Veldink, J.H., Westra, H.-J., Arends, D., Bonder, M.J., Fu, J., Deelen, P., Groen, H.J.M., Smolonska, A. *et al.* (2011) Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.*, **7**, e1002197.
55. Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S. *et al.* (2008) Genetics of gene expression and its effect on disease. *Nature*, **452**, 423–428.
56. Mehta, D., Heim, K., Herder, C., Carstensen, M., Eckstein, G., Schurmann, C., Homuth, G., Nauck, M., Völker, U., Roden, M. *et al.* (2013) Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. *Eur. J. Hum. Genet.*, **21**, 48–54.
57. Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C.C., Taylor, J., Burnett, E., Gut, I., Farrall, M. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
58. Liang, L., Morar, N., Dixon, A.L., Lathrop, G.M., Abecasis, G.R., Moffatt, M.F. and Cookson, W.O.C. (2013) A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res.*, **23**, 716–726.
59. Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D. *et al.* (2007) Population genomics of human gene expression. *Nat. Genet.*, **39**, 1217–1224.
60. Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T.J., Sladek, R. and Majewski, J. (2008) Genome-wide analysis of transcript isoform variation in humans. *Nat. Genet.*, **40**, 225–231.
61. Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M. *et al.* (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*, **325**, 1246–1250.
62. Cusanovich, D.A., Billstrand, C., Zhou, X., Chavarria, C., De Leon, S., Michelini, K., Pai, A.A., Ober, C. and Gilad, Y. (2012) The combination of a genome-wide association study of lymphocyte count and analysis of gene expression data reveals novel asthma candidate genes. *Hum. Mol. Genet.*, **21**, 2111–2123.
63. Grundberg, E., Small, K.S., Hedman, Å.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.-P., Meduri, E., Barrett, A. *et al.* (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.*, **44**, 1084–1089.
64. Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O. and Knight, J.C. (2012) Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.*, **44**, 502–510.
65. Murphy, A., Chu, J.-H., Xu, M., Carey, V.J., Lazarus, R., Liu, A., Szefler, S.J., Strunk, R., Demuth, K., Castro, M. *et al.* (2010) Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. *Hum. Mol. Genet.*, **19**, 4745–4757.
66. Heinzen, E.L., Ge, D., Cronin, K.D., Maia, J.M., Shianna, K.V., Gabriel, W.N., Welsh-Bohmer, K.A., Hulette, C.M., Denny, T.N. and Goldstein, D.B. (2008) Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol.*, **6**, e1.
67. Barreiro, L.B., Tailleux, L., Pai, A.A., Gicquel, B., Marioni, J.C. and Gilad, Y. (2012) Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc. Natl. Acad. Sci. USA*, **109**, 1204–1209.
68. Huang, R.S., Gamazon, E.R., Ziliak, D., Wen, Y., Im, H.K., Zhang, W., Wing, C., Duan, S., Bleibel, W.K., Cox, N.J. *et al.* (2011) Population differences in microRNA expression and biological implications. *RNA Biol.*, **8**, 692–701.