*Human Molecular Genetics*, 2017, Vol. 26, No. 6 1193–1204

**doi: 10.1093/hmg/ddx024**
Advance Access Publication Date: 1 February 2017
Association Studies Article

ASSOCIATION STUDIES ARTICLE

# Genome-wide association of white blood cell counts in Hispanic/Latino Americans: the Hispanic Community Health Study/Study of Latinos

Deepti Jain[1,†], Chani J. Hodonsky[2,†], Ursula M. Schick[3,4,5,†], Jean V. Morrison[1], Sharon Minnerath[6], Lisa Brown[1], Claudia Schurmann[3,4], Yongmei Liu[7], Paul L. Auer[8], Cecelia A. Laurie[1], Kent D. Taylor[9,10], Brian L. Browning[11], George Papanicolaou[12], Sharon R. Browning[1], Ruth J. F. Loos[3,4,13], Kari E. North[2,14], Bharat Thyagarajan[6], Cathy C. Laurie[1], Timothy A. Thornton[1], Tamar Sofer[1] and Alexander P. Reiner[15,*]

[1]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA, [2]Department of Epidemiology, University of North Carolina Gillings School of Public Health, Chapel Hill, NC 27514, USA, [3]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA, [4]The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA, [5]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98195, USA, [6]Department of Lab Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA, [7]Department of Epidemiology and Prevention, School of Medicine, Wake Forest University, Winston-Salem, NC 27101, USA, [8]Department of Biostatistics, Joseph J. Zilber School of Public Health, University of Wisconsin Milwaukee, Milwaukee, WI 53201, USA, [9]Institute for Translational Genomics and Population Sciences at Harbor-UCLA Medical Center, Torrance, CA 90502, USA, [10]Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA, [11]Department of Medicine, University of Washington, Seattle, WA 98195, USA, [12]Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute, Bethesda, MD 20824, USA, [13]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA, [14]Department of Genetics, University of North Carolina, Chapel Hill, NC 27514, USA and [15]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98195, USA

*To whom correspondence should be addressed at: University of Washington, Seattle, WA 98195, USA. Tel: 206 667 2710; Fax: 206 667 4142; Email: apreiner@uw.edu

## Abstract

Circulating white blood cell (WBC) counts (neutrophils, monocytes, lymphocytes, eosinophils, basophils) differ by ethnicity. The genetic factors underlying basal WBC traits in Hispanics/Latinos are unknown. We performed a genome-wide association study of total WBC and differential counts in a large, ethnically diverse US population sample of Hispanics/Latinos

[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.
**Received:** November 11, 2016. **Revised:** January 10, 2017. **Accepted:** January 11, 2017

ascertained by the Hispanic Community Health Study and Study of Latinos (HCHS/SOL). We demonstrate that several previously known WBC-associated genetic loci (e.g. the African Duffy antigen receptor for chemokines null variant for neutrophil count) are generalizable to WBC traits in Hispanics/Latinos. We identified and replicated common and rare germ-line variants at *FLT3* (a gene often somatically mutated in leukemia) associated with monocyte count. The common FLT3 variant rs76428106 has a large allele frequency differential between African and non-African populations. We also identified several novel genetic loci involving or regulating hematopoietic transcription factors (*CEBPE-SLC7A7*, *CEBPA and CRBN-TRNT1*) associated with basophil count. The minor allele of the *CEBPE* variant associated with lower basophil count has been previously associated with Amerindian ancestry and higher risk of acute lymphoblastic leukemia in Hispanics. Together, these data suggest that germline genetic variation affecting transcriptional and signaling pathways that underlie WBC development and lineage specification can contribute to inter-individual as well as ethnic differences in peripheral blood cell counts (normal hematopoiesis) in addition to susceptibility to leukemia (malignant hematopoiesis).

## Introduction

Total white blood cell (WBC) and differential (neutrophil, monocyte, lymphocyte, eosinophil and basophil) counts are indicators of general health and are influenced by inflammatory, immune, allergic and hematologic diseases. During normal hematopoiesis in the bone marrow, WBC production and differentiation are regulated by the coordinated action of hematopoietic growth factor and cytokine signaling, transcriptional regulation and epigenetic modification of lineage-specific genes (1). Dysregulation of these pathways leads to abnormal differentiation and proliferation of immature progenitor cells that underlies the pathogenesis of leukemia and other clonal hematologic disorders (myeloproliferative neoplasms) (2).

Circulating WBC counts are heritable, complex, polygenic traits (3–5) that exhibit a large degree of inter-individual and ethnic variation. Genome-wide association studies (GWAS) have identified approximately 50 genetic loci associated with WBC traits in individuals of European, Asian and African descent (6–12). Total WBC and neutrophil counts tend to be lower on average among individuals of African descent compared to other populations (13). This is partially attributable to the African-derived Duffy antigen receptor for chemokines (DARC) 'null' variant (rs2814778), which is known to confer resistance to *Plasmodium falciparum* malaria infection (14,15).

Hispanics/Latinos in the US are a highly heterogeneous ethnic group with varied proportions of Amerindian, European and African ancestry. Data on WBC trait variability in Hispanic/Latino populations are fairly limited. Despite the greater proportion of African ancestry, Hispanics/Latino on average have been reported to have higher total WBC and neutrophil counts compared to non-Hispanic whites (16,17). Moreover, certain acquired or inherited hematologic conditions such as childhood acute lymphoblastic leukemia (ALL) (18) are more common among Hispanics/Latinos. Whether previously identified or as-yet unidentified shared or ethnicity-specific genetic factors contribute to inter-individual WBC trait variability in Hispanics/Latinos is largely unknown. We therefore performed a GWAS of total WBC count and WBC subtype counts in a large, ethnically diverse US sample ascertained by the Hispanic Community Health Study and Study of Latinos (HCHS/SOL).

## Results

Genome-wide association analysis of total WBC and the absolute number of circulating neutrophils, monocytes, lymphocytes, eosinophils and basophils was performed in 11 809 Hispanic-/Latino- Americans from HCHS/SOL. Genomic inflation factors ranged from 1.008 (basophils) to 1.034 (neutrophils), indicating adequate control of population stratification. Quantile–quantile and Manhattan plots are shown in Figure 1.

Overall, 17 distinct loci, representing 21 combinations of trait-variant associations, were significantly ($P < 5 \times 10^{-8}$) or suggestively ($P \leq 1 \times 10^{-7}$) associated with total WBC and/or WBC subtype in the HCHS/SOL Hispanics (Table 1). Of the 21 trait-variant associations, five were with total WBC, four with neutrophils, one with lymphocytes, five with monocytes, two with eosinophils and four with basophils. Several of these loci (*DARC*, *GATA2*, *HLA-C*, *CSF3-PSMD3* and *SLCO5A1*) were associated with more than one WBC trait.

### Generalization of known WBC-associated loci to HCHS/SOL Hispanics

The genome-wide significant associations in the HCHS/SOL Hispanics/Latinos included several known WBC trait-associated loci (*DARC*, *ITGA4, GATA2, HLA-C*; *CCDC26-GSDMC*; *LPAR1* and *CSF3-PSMD3*) previously identified in GWAS of European, African or Asian descent individuals (Table 1). The most strongly associated variant was the African *DARC* null variant rs2814778 for total WBC and neutrophil count.

We formally tested the generalization of previously reported simple nucleotide polymorphism (SNP) associations with WBC traits to HCHS/SOL Hispanics/Latinos using a directional false discovery rate (FDR)-based procedure described under Methods (Fig. 2; Supplementary Material, Table S1). Of 57 WBC unique trait-SNP associations previously reported in European American, African American or Asian descent samples, 38% or 67% showed evidence of generalization to our Hispanic/Latino discovery sample. On a per trait basis, the replication rate was 1 of 4 loci (25%) for basophils, 6 of 13 loci (46%) for eosinophils, 2 of 2 loci (100%) for lymphocytes, 13 of 16 loci (81%) for monocytes, 4 of 5 loci (80%) for neutrophils and 12 of 17 loci (71%) for total WBC. When stratified by ancestry of the discovery sample, the generalization rate was 72% for European Americans, 86% for African Americans and 73% for Asians.

Our sample size in HCHS/SOL is comparable to that of prior GWAS of WBC traits performed in other populations ($N = 10\,000$–$15\,000$) (6–12). Nonetheless, our ability to detect generalization of SNPs in some instances may be limited by statistical power. To address this question, we further assessed the directional consistency of SNP effect sizes for those SNPs that failed to generalize by generating a genetic score summing all trait-increasing alleles for each HCHS/SOL participant. For total WBC, monocytes, eosinophils and basophils, there were 5, 3, 7 and 3 SNPs, respectively, that failed to generalize. The *P*-value for directionally consistent association of the genetic score was 0.34, 0.005, 0.014 and 0.067, respectively, for total WBC,

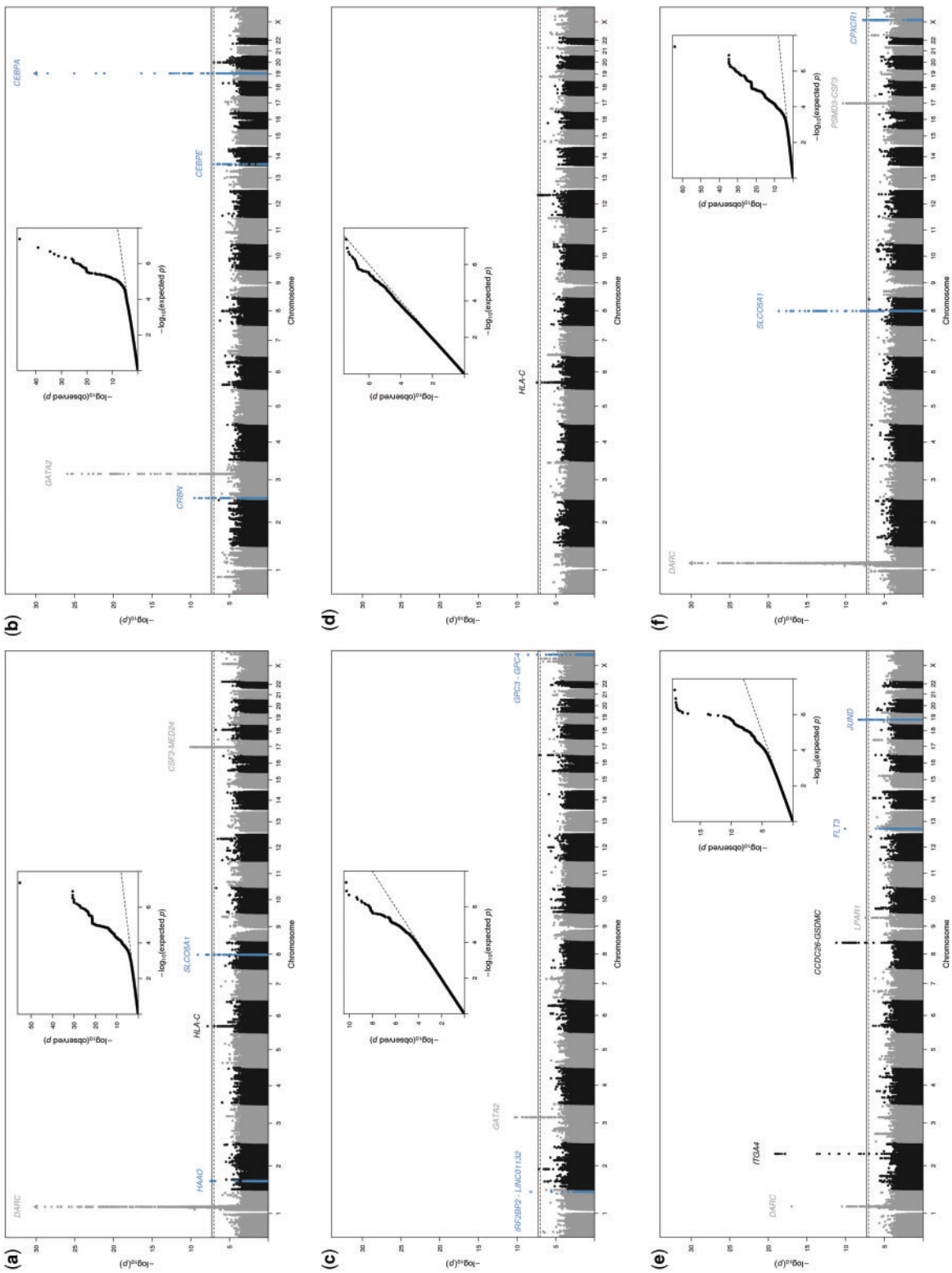


**Figure 1.** Manhattan plots of GWAS for WBC traits in HCHS/SOL. Plots include all variants with MAF >0.01. The x-axis represents physical location of variants on all autosomal chromosomes and the X chromosome; the y-axis is the negative log base-10 P-value. The solid line indicates genome-wide significance (P-value $<5 \times 10^{-8}$), and a dashed line indicates suggestive significance (P-value $<1 \times 10^{-7}$). The inset is a quantile–quantile plot of discovery P-values. Discovery loci are highlighted in blue, and loci with P-values less than the suggestive significance threshold are annotated with the names of the nearest gene(s). **(A)** Total WBC count. **(B)** Basophil count. **(C)** Eosinophil count. **(D)** Lymphocyte count. **(E)** Monocyte count. **(F)** Neutrophil count.

**Table 1.** Loci associated with WBC traits in HCHS/SOL discovery sample

| Trait | SNP | Chr:Position | Locus | Coded/Alt | CAF | N | Beta (SE) | p-value | AFR | AMR | ASN | EUR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total WBC | rs2814778 | 1: 159174683 | *DARC* | T/C | 0.859 | 11,809 | 0.1037 (0.0066) | 5.68E-56 | 0.06 | 0.93 | 1.00 | 1.00 |
| **Total WBC** | **rs114477531** | **2: 43146421** | ***HAAO*** | **T/C** | **0.987** | **11,809** | **0.0909 (0.0165)** | **3.60E-08** | 0.96 | 0.99 | 1.00 | 0.99 |
| Total WBC | rs2524079 | 6: 31242174 | *HLA-C* | G/A | 0.556 | 11,809 | −0.0213 (0.0038) | 1.50E-08 | 0.53 | 0.52 | 0.71 | 0.53 |
| **Total WBC** | **rs2380606** | **8: 70740896** | ***SLCO5A1*** | **T/C** | **0.491** | **11,808** | **0.0240 (0.0039)** | **8.00E-10** | 0.11 | 0.51 | 0.46 | 0.51 |
| Total WBC | rs2227336 | 17: 38174855 | *CSF3-MED24* | T/G | 0.663 | 11,809 | −0.0262 (0.0040) | 8.80E-11 | 0.71 | 0.66 | 0.58 | 0.63 |
| Neutrophil | rs2814778 | 1: 159174683 | *DARC* | T/C | 0.859 | 11,809 | 0.1275 (0.0075) | 5.72E-65 | 0.06 | 0.93 | 1.00 | 1.00 |
| **Neutrophil** | **rs2380606** | **8: 70740896** | ***SLCO5A1*** | **T/C** | **0.491** | **11,808** | **0.0402 (0.0045)** | **2.22E-19** | 0.11 | 0.51 | 0.46 | 0.51 |
| Neutrophil | rs35272691 | 17: 38157841 | *PSMD3-CSF3* | T/C | 0.593 | 11,809 | 0.0297 (0.0045) | 4.51E-11 | 0.90 | 0.58 | 0.60 | 0.64 |
| **Neutrophil** | **rs7882966** | **X: 87982153** | ***CPXCR1*** | **T/C** | **0.955** | **11,797** | **0.0491 (0.0087)** | **1.81E-08** | 0.91 | 0.94 | 1.00 | 0.95 |
| Lymphocyte | rs2249742 | 6: 31240721 | *HLA-C* | C/T | 0.486 | 11,809 | −0.0219 (0.0040) | 3.70E-08 | 0.47 | 0.43 | 0.49 | 0.50 |
| Monocyte | rs201013030 | 2: 182324188 | *ITGA4* | T/C | 0.447 | 11,809 | −0.0132 (0.0015) | 7.72E-20 | NR | NR | NR | NR |
| Monocyte | rs13277237 | 8: 130604563 | *CCDC26-GSDMC* | G/A | 0.459 | 11,808 | 0.0099 (0.0014) | 6.37E-12 | 0.59 | 0.49 | 0.55 | 0.50 |
| Monocyte | rs200243293 | 9: 113945615 | *LPAR1* | T/TG | 0.659 | 11,809 | 0.0085 (0.0016) | 3.88E-08 | 0.81 | 0.62 | 0.86 | 0.53 |
| **Monocyte** | **rs76428106** | **13: 28604007** | ***FLT3*** | **T/C** | **0.990** | **11,809** | **−0.0497 (0.0076)** | **8.17E-11** | 1.00 | 0.99 | 1.00 | 0.99 |
| **Monocyte** | **rs12973608** | **19: 18287220** | ***JUND*** | **A/C** | **0.338** | **11,809** | **−0.0090 (0.0015)** | **4.68E-09** | 0.42 | 0.36 | 0.18 | 0.36 |
| **Eosinophil** | **rs3009958** | **1: 234879890** | ***LINC01132-IRF2BP2*** | **A/G** | **0.015** | **11,789** | **0.0420 (0.0072)** | **6.10E-09** | 0.10 | 0.01 | 0.00 | 0.00 |
| Eosinophil | rs13089722 | 3: 128306757 | *GATA2* | G/A | 0.827 | 11,789 | 0.0147 (0.0022) | 5.40E-11 | 0.95 | 0.87 | 0.64 | 0.89 |
| **Basophil** | **rs1669340** | **3: 3198380** | ***CRBN1*** | **G/T** | **0.316** | **11,789** | **0.0322 (0.0051)** | **3.23E-10** | 0.54 | 0.30 | 0.75 | 0.16 |
| Basophil | rs6782812 | 3: 128317997 | *GATA2* | G/A | 0.173 | 11,789 | −0.0671 (0.0063) | 1.17E-26 | 0.05 | 0.12 | 0.38 | 0.11 |
| **Basophil** | **rs9743723** | **14: 23577198** | ***CEBPE*** | **C/T** | **0.474** | **11,789** | **−0.0249 (0.0049)** | **4.08E-07** | 0.77 | 0.46 | 0.40 | 0.40 |
| **Basophil** | **rs78744187** | **19: 33754548** | ***CEBPA*** | **C/T** | **0.916** | **11,789** | **0.1198 (0.0083)** | **3.87E-47** | 0.98 | 0.91 | 0.98 | 0.93 |

Associations not previously reported are shown in bold.

Coded/Alt, coded, alternative alleles in HCHS/SOL; CAF, coded allele frequency; SE, standard error; 1000 Genomes super-population CAF: AFR, African; AMR, admixed American; ASN, Asian; EUR, European; NR, not reported.

monocyte, eosinophil and basophil counts (Supplementary Material, Table S2). These results suggest that, at least for monocytes, eosinophils and basophils, low power may contribute to lack of generalization. Failure to generalize might also occur because of differences in allele frequency or linkage disequilibrium (LD) patterns for the index SNP in HCHS/SOL compared to the original population.

## Discovery and replication of novel WBC-associated loci

Ten of the genome-wide significant WBC trait-locus associations in HCHS/SOL were not previously reported through GWAS of quantitative WBC traits (Table 1). These included *LINC01132—LOC101927851* rs3009958 for eosinophils; *HAAO* rs114477531 for total WBC; *CRBN* rs1669340 for basophils; *SLCO5A1* rs2380606 for total WBC and neutrophils; *FLT3* rs76428106 for monocytes; *JUND* rs12973608 for monocytes; *CEBPA* rs78744187 for basophils and *CPXCR1* rs7882966 for neutrophils. In addition to the 10 novel genome-wide significant loci, a SNP on chr14q11 (rs9743723) had a suggestive association with basophils ($P = 4 \times 10^{-7}$). The lead SNP rs9743723 is located downstream of *CEBPE*, which encodes a hematopoietic transcription factor involved in terminal granulocyte differentiation. Other *CEBPE* variants have been associated with risk of leukemia.

To assess the presence of secondary, independent association signals at any of our genome-wide significant regions, we carried out conditional analyses adjusting for the effect of the lead variant at each genome-significant WBC trait locus in HCHS/SOL. The only locus that showed evidence of independent association signals ($P < 5 \times 10^{-8}$ following conditional analysis) was *FLT3*, where there was evidence of two independent genome-wide association signals for monocyte count. The lead

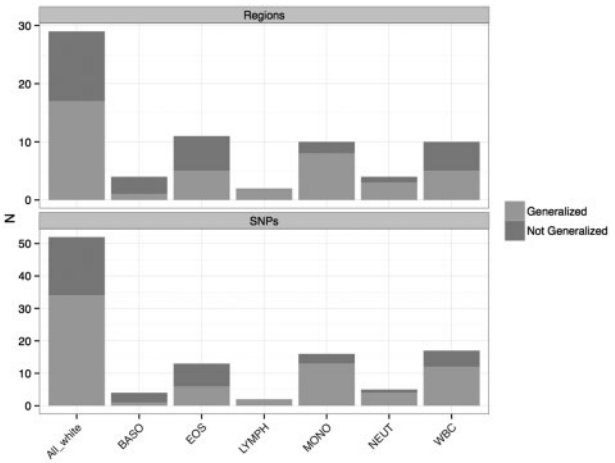


**Figure 2.** Generalization of previously reported WBC trait loci to HCHS/SOL. The top panel contains generalization results for all previously reported loci by trait (x-axis). The bottom panel contains generalization results for all previously reported index variants by trait. The total number of previously reported loci or index variants per trait (y-axis) is shaded dark gray, and the number of generalized loci or index variants per trait ($r<0.05$) is shaded light gray.

variant *FLT3* rs76428106 C allele (MAF = 1%) was associated with higher monocyte count ($P = 8.17 \times 10^{-11}$). After conditioning on the lead variant, the *FLT3* rs7327579 A allele (MAF = 48%) was associated with higher monocyte count ($P = 1.24 \times 10^{-8}$) (Fig. 3). Prior to conditional analysis, the *P*-value for monocyte association for rs7327579 was $3.26 \times 10^{-6}$.

On the basis of the HCHS/SOL discovery-stage results and conditional analysis, we selected 11 variants for follow-up/replication testing in our Hispanic/Latino validation sample

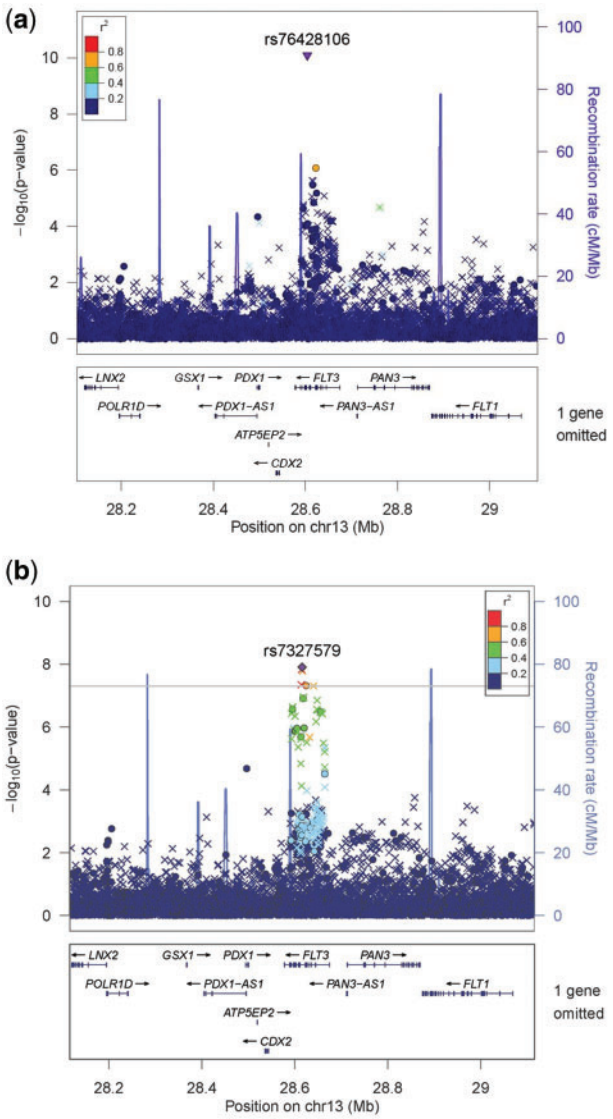


**Figure 3.** LocusZoom plots showing two independent association signals for monocyte count at the *FLT3* locus. The Panel (a) contains a LocusZoom plot of the FLT3 locus centered on our top variant, rs76428106 (imputed), indicated by a purple triangle. The Panel (b) contains a LocusZoom plot of the FLT3 locus centered on rs7327579, the top variant after conditional analysis on rs76428106. rs7327579 is genotyped and hence represented by a purple diamond. The LD estimates derived from the HCHS/SOL study samples with respect to the top variant and the other variants in the window are color-coded by correlation category according to the scale in the upper right of each panel. Imputed variants are denoted by an x, and genotyped variants are denoted by a filled circle. Recombination hotspots from HapMap are indicated by vertical blue peaks. Genes within the region of interest are listed by chromosomal position beneath the x-axis. The horizontal line indicates the significance threshold P-value, $5 \times 10^{-8}$.

($N = 7\ 200$) from the Women's Health Initiative (WHI) SNP Health Association Resource (SHARe) project, the Multi-Ethnic Study of Atherosclerosis (MESA) cohort and the Mount Sinai BioMe Biobank. These included nine novel genome-wide-significant associations (rs2380606 was associated with both neutrophil count and total WBC), the conditionally independent *FLT3* rs7327579 variant for monocyte count, and the suggestive basophil-associated variant *CEBPE* rs9743723. Nine of ten total variants (except *CPXCR1* rs7882966) were available for testing the 11 associations. Of ten associations tested, four were replicated ($P < 0.005$ with directional consistency) in the Hispanic/Latino validation sample (Table 2): *FLT3* rs7327579 for monocytes, *CEBPA* rs78744187 for basophils, *CEBPE* rs9743723 for basophils and *CRBN* rs1669340 for basophils and a fifth variant *FLT3* rs76428106 for monocytes had a suggestive replication ($P < 0.009$).

While this paper was under review, *FLT3* rs76428106 and *CEBPA* rs78744187 were reported to be associated with monocyte-related and basophil-related traits, respectively, in Europeans in a large UK Biobank meta-analysis (19). In addition, the minor allele of *FLT3* rs76428106 was significantly associated with higher total WBC and the minor allele of *CEBPA* rs78744187 was significantly associated with several red blood cell (RBC) traits [higher RBC count and lower mean corpuscular volume (MCV) and mean corpuscular hemoglobin (MCH)]. In HCHS/SOL, although none of these additional blood cell trait associations reached genome-wide significance, we confirmed that *FLT3* rs76428106 was nominally associated with higher total ln(WBC) (beta = 0.043, SE = 0.20; $P = 0.03$), while *CEBPA* rs78744187 was nominally associated with RBC count (beta = 0.042, SE = 0.0085; $P = 6.6 \times 10^{-7}$), MCV (beta = −0.40, SE = 0.13; $P = 0.0024$) and MCH (beta = −0.14, SE = 0.05; $P = 0.0031$).

## Functional annotation and overlap of newly identified *FLT3*, *CEBPA*, *CEBPE* and *CRBN-TRNT1* genetic variants with regulatory elements

*FLT3* rs76428106 is in LD ($r^2 = 0.6$) with rs79490353, which overlaps a genomic element enriched for H3K4me1 and H3K27ac enhancer histone marks in monocytes (20) and also for H3K4me1, H3K27ac and H3K4me3 in lymphoblastoid cells (GM12878) and primary B cells (21). In GM12878 lymphoblastoid cells, the genomic element harboring rs79490353 also overlaps a ChIP-Seq (chromatin immunoprecipitation followed by sequencing) peak for the myeloid transcription factor PU.1 (21). RS7327579 is located within an open chromatin region of a *FLT3* intron enriched with the enhancer histone mark H3K4me1 that is found predominantly in monocytes and dendritic cells (20). Interestingly, the DNase site is hypersensitive in two out of four BLUEPRINT monocyte donors, further supporting a potential role for allele-specific differences in *FLT3* gene regulation by this variant. In Europeans and Amerindians, rs7327579 is also in strong LD ($r^2 = 0.8$) with a common *FLT3* missense variant rs1933437 (p.Thr227Met) located in the second Ig-like domain of the extracellular ligand-binding region. This amino acid substitution is predicted to be likely damaging (22).

The *CEBPA* rs78744187 lead variant is a strong functional candidate due to its location in a K562 myeloid leukemia cell line, CD34+ cell and monocyte DNaseI site and active enhancer region that binds several hematopoietic transcription factors (CEBPβ, GATA-2 and TAL-1) in a blood cell-specific manner (20,22). Another myeloid-specific regulatory element (enriched for H3K4me1, H3K4me3 and H3K27ac marks in monocytes, macrophages and neutrophils) harbors two additional LD proxy variants, rs958483 and rs73926283.

The lead SNP *CEBPE* rs9743723 is in LD with several candidate functional SNPs located within promoter or enhancer histone marks and some of these variants have been suggested to influence CEBPE expression (23,24). For example, the LD proxy variant rs10143875 is located within a transcribed *CEBPE* enhancer and is bound *in vitro* by several transcription factors including PU.1 and Jun in granulocytes and monocytes (20,22). It is noteworthy that rs10143875 putatively alters binding motifs for BCL6B and STAT5A; the latter transcription factor has been

**Table 2.** Replication and meta-analysis of 10 discovery WBC associations in three independent Hispanic/Latino cohorts

| | | | | BioMe | | | MESA | | | WHI | | | Meta-analysis | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trait | Locus | rsid | Coded/Alt | N | Beta (SE) | P-value | N | Beta (SE) | P-value | N | Beta (SE) | P-value | N | Beta (SE) | P-value | $^{b}I^2$ |
| Total WBC | *HAAO* | rs114477531 | T/C | 2855 | 0.019 (0.087) | 0.83 | 782 | 0.036 (0.067) | 0.59 | 3546 | −0.074 (0.029) | 0.015 | 7183 | −0.050 (0.025) | 0.04 | 0 |
| Total WBC | *SLCO5A1* | rs2380606 | T/C | 2855 | 0.014 (0.027) | 0.60 | 782 | −0.017 (0.013) | 0.22 | 3546 | 0.0082 (0.0063) | 0.20 | 7183 | 0.0042 (0.0056) | 0.45 | 0 |
| Neutrophil | *SLCO5A1* | rs2380606 | T/C | 2230 | 0.0070 (0.0030) | 0.82 | 782 | −0.022 (0.019) | 0.24 | 1205 | 0.028(0.015) | 0.06 | 4217 | 0.0079 (0.011) | 0.47 | 7.2 |
| **Monocyte** | ***FLT3*** | **rs76428106** | **T/C** | **2230** | **−0.334 (0.160)** | **0.04** | **782** | **−0.081 (0.048)** | **0.10** | **1205** | **−0.252 (0.147)** | **0.09** | **4217** | **−0.1152 (0.044)** | **0.009** | **0** |
| **Monocyte** | ***FLT3*** | **rs7327579**[a] | **G/A** | **2230** | **0.043 (0.030)** | **0.16** | **782** | **0.018 (0.006)** | **0.005** | **1205** | **0.022 (0.014)** | **0.13** | **4217** | **0.0194 (0.0057)** | **7.2E-04** | **0** |
| Monocyte | *JUND* | rs12973608 | A/C | 2230 | −0.052 (0.031) | 0.87 | 782 | −0.007 (0.006) | 0.23 | 1205 | 0.0017 (0.013) | 0.90 | 4217 | −0.0069 (0.0052) | 0.19 | 0 |
| Eosinophil | *LINC01132 -IRF2BP2* | rs3009958 | A/G | 2122 | −0.011 (0.095) | 0.91 | 782 | −0.034 (0.030) | 0.25 | 1205 | 0.062 (0.031) | 0.04 | 4109 | 0.012 (0.021) | 0.58 | 22.3 |
| **Basophil** | ***CRBN1*** | **rs1669340** | **G/T** | **2215** | **0.164 (0.031)** | **1.7E-07** | **782** | **0.040 (0.056)** | **0.48** | **1205** | **0.051(0.020)** | **0.01** | **4202** | **0.081 (0.016)** | **6.6E-07** | **59.3** |
| **Basophil** | ***CEBPE*** | **rs9743723** | **C/T** | **2230** | **−0.101 (0.031)** | **1.3E-03** | **782** | **−0.041 (0.055)** | **0.45** | **1205** | **−0.085 (0.026)** | **0.001** | **4210** | **−0.086 (0.019)** | **5.9E-06** | **0** |
| **Basophil** | ***CEBPA*** | **rs78744187** | **C/T** | **2230** | **0.223 (0.058)** | **1.1E-04** | **782** | **0.465 (0.127)** | **3.0E-4** | **1205** | **0.271 (0.070)** | **1.1E-04** | **4210** | **0.267 (0.042)** | **2.2E-10** | **0** |

Replicated associations shown in bold.

Coded/Alt, coded, alternative alleles in HCHS/SOL; SE, standard error.

[a]rs7327579 had P-value 1.24E-8 in a joint analysis with rs76428106. Tested independently, not jointly with rs76428106, in replication cohorts.

[b]I-squared represents heterogeneity of effect as a percent between 0 and 100. I-squared >80 indicates significant heterogeneity.

implicated in basophil and mast cell lineage specification and differentiation (25). Another proxy SNP, rs2239635, was reported to disrupt binding of the hematopoietic transcriptional repressor Ikaros, which is a negative regulator of basophil production (26).

The chromosome 3p26 association signal for higher basophil count encompasses several non-coding variants which are cis-eQTL for *CRBN* and *TRNT1* in whole blood. The lead *CRBN* SNP rs1669340 and several proxy variants are located within blood cell epigenomic promoter or enhancer marks (20,22).

## Discussion

In a GWAS of nearly 12 000 US Hispanics/Latinos, we discovered and replicated several WBC trait loci. These include two independently associated *FLT3* signals for monocyte count, two CCAAT/enhancer binding proteins (*CEBPE* and *CEBPA*) for basophil count and *CRBN* for basophil count. The two distinct *FLT3* variants rs76428106 and rs7327579 are low-frequency and common, respectively, and have different allele frequencies in different ancestry groups. We also demonstrated the generalizability to Hispanics/Latinos of the majority of WBC trait loci previously identified in GWAS of European, African or Asian ancestry. The new *FLT3*, *CEBPE*, *CEBPA* and *CRBN* loci suggest that hematopoiesis-related genes that are grossly altered in malignant hematopoiesis can additionally harbor genetic variants with smaller effect magnitudes that influence circulating levels of WBC subtypes in individuals with no known hematological cancers.

### *FLT3* variants associated with monocyte count

The lead *FLT3* variant rs76428106 (associated with higher monocyte count) is intronic and has an allele frequency of ~1% in European and Amerindian populations but is monomorphic in African and Asian 1000 Genomes reference populations. After conditioning on the lead *FLT3* variant rs76428106, a second *FLT3* variant (rs7327579) was independently associated with higher monocyte count. rs7327579 is common (MAF = 48%), but has a large allele frequency differential between African (A allele frequency = 14%) and non-African populations (A allele frequency 54% in Amerindian, 53% in European, 64% in South Asian and 70% in East Asian 1000 Genomes superpopulations).

FLT3 is a receptor tyrosine kinase that regulates early hematopoiesis (27) and is frequently mutated in hematologic malignancies and is an important prognostic factor and therapeutic target for acute myeloid leukemia (AML) (28,29). FLT3 is expressed not only on early myeloid progenitors, but also on monocytes and thereby promotes monocyte/macrophage proliferation and differentiation as well as development and activation of monocyte-derived dendritic cells (30,31). The two monocyte count-associated *FLT3* variants identified in individuals from HCHS/SOL with no know hematological cancers are distinct from the well-characterized activating *FLT3* somatic mutations frequently observed in AML. The *FLT3* leukemic mutations generally consist of internal tandem duplications and point mutations involving, respectively, the FLT3 juxtamembrane domain and tyrosine kinase domain, and lead to constitutive activation of the FLT3 receptor and dysregulated downstream signaling pathways and ligand-independent myeloid proliferation (29). The FLT3 monocyte count-associated variants identified in HCHS/SOL likely have subtler effects on later stages of granulocyte/monocyte lineage specification.

Given the poorer prognosis of AML among African Americans compared to whites (32), additional study of *FLT3* variants in AML outcomes may be warranted.

### *CEBPA* and basophil count

The basophil association signal on chr19q13.11 is located approximately 30 kb downstream from *CEBPA*, which encodes C/EBPα, a transcription factor that plays an essential role in myeloid differentiation and specification of neutrophil, monocyte and basophil lineage fates (33). The timing of expression of *GATA2* and *CEBPA* are important for myeloid lineage fate, including the production and differentiation of basophils and mast cells (34). The minor allele of the lead SNP associated with lower basophil count (rs78744187, T allele frequency = 8.4%) is ~3-fold more common in Amerindians and Europeans than Africans and Asians.

Highly penetrant germline mutations and acquired somatic mutations of *CEBPA* each contribute to abnormal WBC maturation and the development of AML (25,26). Since the basophil-associated candidate functional SNP rs73926283 disrupts a binding site for the ZFX hematopoietic proto-oncogene, further assessment of this basophil count-associated variant may be warranted in the context of leukemia disease progression or treatment resistance (35).

### *CEBPE* and basophil count

The region on chr14q11.2 associated with basophil count is localized to a 13 kb LD block downstream of *CEBPE*, encoding CCAAT/enhancer-binding protein ϵ, another CEBP family transcription factor involved in myelopoiesis and terminal granulocyte differentiation (36–38). The minor allele of the basophil-lowering HCHS/SOL index SNP rs9743723 (C, allele frequency = 47%) is also in moderate LD with a set of *CEBPE* SNPs increased susceptibility to childhood ALL (39,40). The higher incidence of ALL and poorer outcomes among Hispanic children have been attributed in part to genetic risk factors associated with Amerindian ancestry (41,42). It is interesting to note that the myeloid enhancer containing rs22239630 is bound by Pol II in NB4 cells (an acute promyelocytic leukemia cell line). The same SNP is predicted to disrupt a putative binding site for ZFX, a transcription factor that maintains a stem/progenitor-like immature phenotype and proliferative capacity of leukemia cells in AML, CML and T-ALL (43). Given the role of CEBPϵ as a suppressor of myeloid leukemogenesis, the newly identified basophil-associated *CEBPE* variant may have additional clinical and therapeutic implications for hematologic cancers (44,45).

### *CRBN* and basophil count

The chromosome 3p26 association signal for higher basophil count encompasses several non-coding variants of *CRBN* and *TRNT1*. Loss-of-function mutations in *TRNT1* result in sideroblastic anemia with B-cell immunodeficiency, periodic fevers and developmental delay (46). CRBN (cereblon), a component of the substrate receptor for an E3 ubiquitin ligase, was recently identified as the molecular target of lenalidomide (LEN), a thalidomide derivative and immunomodulatory drug used to treat hematologic malignancies such as multiple myeloma and 5q-deletion-associated myelodysplastic syndrome (47,48). LEN inhibits ubiquitination of endogenous CRBN substrates and also alters ligase substrate specificity to target new proteins for degradation including IKZF1, a transcription factor important for basophil development and a tumor suppressor for leukemia (49).

### Other WBC loci that generalize to Hispanics/Latinos and implications for WBC/immune-related diseases

Other genome-wide significant associations in the HCHS/SOL included several known WBC trait-associated loci previously identified in GWAS of European-, African- or Asian-descent individuals containing genes involved in WBC production, migration or clearance from the circulation (see Table 1). Several of these loci may have implications for ethnic or racial disparities in chronic disease health outcomes. For example, the DARC-null rs2814778 genotype has been under selection among Africans as a receptor for *Plasmodium vivax* malaria and is also a major determinant for ethnic neutropenia, which could have impact on the pathogenesis of diseases such as HIV or minority participant eligibility in cancer clinical trials (due to exclusion on the basis of low blood counts) (13). Other WBC-associated variants are associated with inflammatory and autoimmune diseases such as *GSDMC* with inflammatory bowel disease (50), *CSF3-PSMD3* with asthma (51), and HLA-C with psoriasis (52), rheumatoid arthritis (53) and Crohn's disease (50). The *GATA2* variant associated with basophil and eosinophil counts in European Americans and Hispanics/Latinos may have implications for the occurrence of hematopoietic disorders (54) as well as allergic diseases such as asthma (55), which varies in prevalence and morbidity among Hispanic/Latino ethnic subgroups in the United States (56).

## Conclusions

In summary, we provide evidence for considerable shared genetic architecture of WBC traits between Hispanics/Latinos and other ethnic groups. We identified several novel WBC trait loci, including variants of *FLT3* associated with monocyte count, and variants of three known regulators of granulocyte and basophil differentiation (*CEBPE*, *CEBPA* and *CRBN*) associated with basophil count. All four of these hematopoiesis-related genes are often grossly altered in hematologic malignancies. Thus, genetic variants with subtler regulatory effects on the same genes may influence circulating levels of WBC subtypes during normal hematopoiesis. Further studies are warranted to assess whether any of these basophil-associated genetic variants are associated with disease susceptibility or outcomes related to leukemia and myeloproliferative neoplasms.

## Materials and Methods

### HCHS/SOL population

The HCHS/SOL is a community-based cohort study of 16 415 self-identified Hispanic/Latino persons aged 18–74 years selected from households in predefined census-block groups from four US field centers (Chicago, Miami, the Bronx and San Diego). Participants self-identified as having a Hispanic/Latino background; the largest groups were Central American ($n = 1\,730$), Cuban ($n = 2\,348$), Dominican ($n = 1\,460$), Mexican ($n = 6\,471$), Puerto Rican ($n = 2\,728$) and South American ($n = 1\,068$). The sample design and cohort selection have been previously described (57). HCHS/SOL participants were recruited between 2008 and 2011 and underwent a baseline clinical examination (58) including biological, behavioral and sociodemographic

assessments. The study was approved by the institutional review boards at each field center, where all subjects gave written informed consent.

### Measurement of WBC and exclusion criteria in HCHS/SOL

Total WBC and differential counts were measured in EDTA whole blood obtained at the baseline examination using a Sysmex XE-2100 instrument (Sysmex America) at the University of Minnesota according to national and international standards and procedures. Individuals pregnant at the time of blood draw; those with >5% circulating blasts or immature cells, end-stage renal disease or any hematologic malignancy; and those undergoing chemotherapy for solid tumors were excluded from our analyses.

### Genotyping, imputation and quality control in HCHS/SOL

Consenting HCHS/SOL subjects were genotyped at Illumina on the HCHS/SOL custom 15041502 B3 array. The custom array comprised the Illumina Omni 2.5M array (HumanOmni2.5-8v.1-1) plus ∼150 000 custom SNPs including ancestry-informative markers, known GWAS hits and drug absorption, distribution, metabolism and excretion (ADME) markers, and SNPs selected from the CLM (Colombian in Medellin, Colombia), MXL (Mexican Ancestry in Los Angeles, California) and PUR (Puerto Rican in Puerto Rico) samples in the 1000 Genomes phase 1 data to capture a greater amount of Amerindian genetic variation.

We applied standardized quality-assurance and quality-control (QA/QC) methods (59) to generate recommended SNP- and sample-level quality filters. Samples were checked for sex discrepancies, gross chromosomal anomalies, relatedness and population structure, missing call rates, batch effects and duplicate-sample discordance. After excluding participants who were ineligible or missing phenotype or genotype data, 11 809 study participants were available for analysis. SNPs were checked for Hardy–Weinberg equilibrium, minor allele frequency (MAF), duplicate-probe discordance, Mendelian errors and missing call rate. A total of 2 232 944 SNPs passed filters for both quality and informativeness (polymorphic and unduplicated) and were carried forward for imputation and downstream association analyses.

Genome-wide imputation in HCHS/SOL was carried out using the full, cosmopolitan 1000 Genomes Project phase 1 reference panel ($n = 1\,092$) (60), as previously described (61). Briefly, genotypes were first pre-phased with SHAPEIT2 (v.2.r644) and then imputed with IMPUTE2 (v.2.3.0) (62,63). Overall imputation quality was assessed by calculating 'oevar' (the ratio of the observed variance of imputed dosages to the expected binomial variance) using the MaCH imputation software (64) and by examination of the distribution of imputation quality metrics from the IMPUTE2 internal masking experiments. We performed downstream association analyses on observed variants passing quality filters and all imputed variants (a total of 27 887 661 variants). Results were then filtered on the basis of imputation quality (oevar > 0.3) and allele frequency (MAF > 1%).

### Linear mixed-effect model for association testing in HCHS/SOL

We analyzed WBC phenotypes by using linear mixed-effect models (LMMs) to account for the correlations due to genetic relatedness (kinship), shared household and block group between individuals. All analyses were adjusted for sex, age, five principal components (PCs), recruitment center, smoking status, log of sampling weights (to prevent potential selection bias due to the sampling scheme) and genetic-analysis group (a six-level categorical variable derived from genetic data and self-identified background) (61). To approximate normal distribution of the model's residuals, the outcomes were either log-transformed (WBC, lymphocytes) or log transformed after addition of the constant 1 (neutrophils, monocytes, eosinophils, basophils). For basophils, we also performed a sensitivity analyses due to the large numbers of zero values due to rounding. In the sensitivity analysis, we dichotomized basophil count to 0 vs. non-zero, and we applied a score test based on a logistic mixed model, with the same fixed and random effects as before (65).

### Replication of discovery loci in independent Hispanic/Latino samples

To replicate association findings in Hispanic/Latino samples, we used 1000 Genomes imputed GWAS data available in three additional Hispanics/Latinos samples, including up to 3 454 from the WHI SHARe project (66); 782 from the MESA cohort (67,68) and 2 854 from Mount Sinai BioMe Biobank (69). WHI-SHARe and MESA participants were genotyped with the Affymetrix 6.0 chip, and imputation was performed with MaCH (64). BioMe participants were genotyped with the Illumina HumanOmniExpressExome-8 v.1.0 chip, and imputation was performed with IMPUTE2 (62,63) in 1000 Genomes phase 1 data (March 2012 v.3). Association testing for typed or imputed SNPs was performed by linear regression of log-transformed WBC count adjusted for age, sex and PCs. Meta-analysis of results from the three replication cohorts for total WBC and WBC subtypes was performed with the inverse-variance-weighted method implemented in METAL (70). To declare significance for replicated WBC loci, we required that (a) the directionality of effect was similar across the discovery and replication phases and (b) the replication $P$-value met the Bonferroni corrected criteria for multiple testing ($P$-value $< 0.05/10 = 0.005$).

### Generalization in HCHS/SOL

We performed generalization analysis for WBC-associated SNPs previously reported in GWASs of other populations, including those of European, African and Japanese ancestry (2,6–11). In testing for generalization we controlled the directional FDR of the generalization null hypotheses at a threshold of 0.05 (71, in press). The generalization null hypothesis states that the effect does not exist in either the discovery study nor in HCHS/SOL and is rejected if there is enough evidence that a SNP affects the outcome, with the same direction of effect, in both the discovery study and HCHS/SOL. A SNP was generalized if its $r$ value $< 0.05$.

### Functional annotation of discovery loci

We interrogated the WBC-associated loci to determine whether the identified non-coding SNPs and indels and correlated proxy

variants ($r^2 \geq 0.5$, calculated in the HCHS/SOL discovery population) were positioned within predicted regulatory regions, namely enhancers and promoters and the nearest biologically plausible gene or genes. These regulatory regions were identified on the basis of the enrichment of various histone-modification and ChIP-seq signals in WBCs (granulocytes, lymphocytes, monocytes, dendritic cells) and bone marrow precursor cells from the Blueprint project (20). A genomic element enriched with the histone H3K4me1 signal was categorized as an enhancer, whereas a genomic element enriched with the histone H3K4me3 signal was categorized as a promoter. SNPs or indels belonging to either promoter or enhancer categories that overlap a DNase I hypersensitive site (a general biochemical feature of regulatory regions) in blood or bone marrow-derived cells were prioritized as putatively functional variants. We also reported whether a given SNP overlaps with any transcription factor ChIP-seq peaks (72). Moreover, because related cell types can share similar regulatory regions, we additionally reported supplementary annotation by using data on other myeloid lineage cells and GM12878 lymphoblastoid cells (72). To identify the motifs disrupted by alleles, we utilized HaploReg (v4) (21) and the JASPAR motif database (73).

## Supplementary Material

Supplementary Material is available at *HMG* online.

## Acknowledgements

We thank the participants and staff of the HCHS/SOL study for their contributions to this study.

*Conflict of Interest statement*. None declared.

## Funding

The baseline examination of HCHS/SOL was carried out as a collaborative study supported by contracts from the National Heart, Lung and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236) and San Diego State University (N01-HC65237). The following Institutes/Centers/Offices contributed to the first phase of HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research (NIDCR), National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements. The Genetic Analysis Center at Washington University was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Additional analysis support was provided by 1R01DK101855-01 and 13GRNT16490017. Genotyping efforts were supported by NHLBI HSN 26220/20054C, NCATS CTSI grant UL1TR000123 and NIDDK Diabetes Research Center (DRC) grant DK063491. This manuscript has been reviewed by the HCHS/SOL Publications Committee for scientific content and consistency of data interpretation with previous HCHS/SOL publications. This research was supported in part by the SOL (Study of Latinos) Grant—a sub-award issued under the Prime Contract No. HHSB268201200054C between HHS, NIH, National Heart, Lung and Blood Institute and Illumina, Inc. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124 and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. SRB was supported by R01-GM110068. The Mount Sinai BioMe Biobank Program is supported by The Andrea and Charles Bronfman Philanthropies. Analyses of BioMe data was supported in part through the computational resources and staff expertize provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai. The UW genetic analysis center was supported by a contract from NHLBI for HCHS/SOL (HHSN268201300005C). A.P.R. was supported by NHLBI (R01HL129132). D.E.B. is supported by NIDDK (K08DK093705) and the Doris Duke Charitable Foundation, Charles H. Hood Foundation, American Society of Hematology, Burroughs Wellcome Fund and Cooley's Anemia Foundation.

## References

1. Orkin, S.H. and Zon, L.I. (2008) Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*, **132**, 631–644.
2. Sive, J.I. and Gottgens, B. (2014) Transcriptional network control of normal and leukaemic haematopoiesis. *Exp. Cell Res.*, **329**, 255–264.
3. Evans, D.M., Frazer, I.H. and Martin, N.G. (1999) Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res*, **2**, 250–257.
4. Pilia, G., Chen, W.M., Scuteri, A., Orru, 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.
5. Whitfield, J.B. and Martin, N.G. (1985) Genetic and environmental influences on the size and number of cells in the blood. *Genet. Epidemiol.*, **2**, 133–144.
6. 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.
7. Ferreira, M.A., Hottenga, J.J., Warrington, N.M., Medland, S.E., Willemsen, G., Lawrence, R.W., Gordon, S., de Geus, E.J., Henders, A.K., Smit, J.H. *et al.* (2009) Sequence variants in three loci influence monocyte counts and erythrocyte volume. *Am. J. Hum. Genet.*, **85**, 745–749.
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. 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.
10. 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.
11. 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.

12. 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.
13. Thobakgale, C.F. and Ndung'u, T. (2014) Neutrophil counts in persons of African origin. *Curr. Opin. Hematol.*, **21**, 50–57.
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. Reich, D., Nalls, M.A., Kao, W.H., 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.
16. Grann, V.R., Bowman, N., Joseph, C., Wei, Y., Horwitz, M.S., Jacobson, J.S., Santella, R.P. and Hershman, D.L. (2008) Neutropenia in 6 ethnic groups from the Caribbean and the U.S. *Cancer*, **113**, 854–860.
17. Hsieh, M.M., Everhart, J.E., Byrd-Holt, D.D., Tisdale, J.F. and Rodgers, G.P. (2007) Prevalence of neutropenia in the U.S. population: age, sex, smoking status, and ethnic differences. *Ann. Intern. Med.*, **146**, 486–492.
18. Yamamoto, J.F. and Goodman, M.T. (2008) Patterns of leukemia incidence in the United States by subtype and demographic characteristics, 1997-2002. *Cancer Causes Control*, **19**, 379–390.
19. Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A. *et al.* (2016) The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*, **167**, 1415–1429. e1419.
20. Adams, D., Altucci, L., Antonarakis, S.E., Ballesteros, J., Beck, S., Bird, A., Bock, C., Boehm, B., Campo, E., Caricasole, A. *et al.* (2012) BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol.*, **30**, 224–226.
21. Ward, L.D. and Kellis, M. (2016) HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.*, **44**, D877–D881.
22. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
23. Ryoo, H., Kong, M., Kim, Y. and Lee, C. (2013) Identification of functional nucleotide and haplotype variants in the promoter of the CEBPE gene. *J. Hum. Genet.*, **58**, 600–603.
24. Wiemels, J.L., de Smith, A.J., Xiao, J., Lee, S.T., Muench, M.O., Fomin, M.E., Zhou, M., Hansen, H.M., Termuhlen, A., Metayer, C. *et al.* (2016) A functional polymorphism in the CEBPE gene promoter influences acute lymphoblastic leukemia risk through interaction with the hematopoietic transcription factor Ikaros. *Leukemia*, **30**, 1194–1197.
25. Li, Y., Qi, X., Liu, B. and Huang, H. (2015) The STAT5-GATA2 pathway is critical in basophil and mast cell differentiation and maintenance. *J. Immunol.*, **194**, 4328–4338.
26. Rao, K.N., Smuda, C., Gregory, G.D., Min, B. and Brown, M.A. (2013) Ikaros limits basophil development by suppressing C/EBP-alpha expression. *Blood*, **122**, 2572–2581.
27. Rusten, L.S., Lyman, S.D., Veiby, O.P. and Jacobsen, S.E. (1996) The FLT3 ligand is a direct and potent stimulator of the growth of primitive and committed human CD34+ bone marrow progenitor cells in vitro. *Blood*, **87**, 1317–1325.
28. Kottaridis, P.D., Gale, R.E., Frew, M.E., Harrison, G., Langabeer, S.E., Belton, A.A., Walker, H., Wheatley, K., Bowen, D.T., Burnett, A.K. *et al.* (2001) The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood*, **98**, 1752–1759.
29. Pawar, R., Bali, O.P., Malhotra, B.K. and Lamba, G. (2014) Recent advances and novel agents for FLT3 mutated acute myeloid leukemia. *Stem Cell Investig.*, **1**, 7.
30. Kim, S.W., Choi, S.M., Choo, Y.S., Kim, I.K., Song, B.W. and Kim, H.S. (2015) Flt3 ligand induces monocyte proliferation and enhances the function of monocyte-derived dendritic cells in vitro. *J. Cell Physiol.*, **230**, 1740–1749.
31. Liu, K., Victora, G.D., Schwickert, T.A., Guermonprez, P., Meredith, M.M., Yao, K., Chu, F.F., Randolph, G.J., Rudensky, A.Y. and Nussenzweig, M. (2009) In vivo analysis of dendritic cell development and homeostasis. *Science*, **324**, 392–397.
32. Children's Oncology, G., Aplenc, R., Alonzo, T.A., Gerbing, R.B., Smith, F.O., Meshinchi, S., Ross, J.A., Perentesis, J., Woods, W.G., Lange, B.J. *et al.* (2006) Ethnicity and survival in childhood acute myeloid leukemia: a report from the Children's Oncology Group. *Blood*, **108**, 74–80.
33. Friedman, A.D. (2015) C/EBPalpha in normal and malignant myelopoiesis. *Int. J. Hematol.*, **101**, 330–341.
34. Iwasaki, H., Mizuno, S., Arinobu, Y., Ozawa, H., Mori, Y., Shigematsu, H., Takatsu, K., Tenen, D.G. and Akashi, K. (2006) The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. *Genes Dev.*, **20**, 3010–3021.
35. Wu, J., Wei, B., Wang, Q., Ding, Y., Deng, Z., Lu, X. and Li, Y. (2016) ZFX facilitates cell proliferation and imatinib resistance in chronic myeloid leukemia cells. *Cell Biochem. Biophys.*, **74**, 277–283.
36. Chumakov, A.M., Grillier, I., Chumakova, E., Chih, D., Slater, J. and Koeffler, H.P. (1997) Cloning of the novel human myeloid-cell-specific C/EBP-epsilon transcription factor. *Mol. Cell Biol.*, **17**, 1375–1386.
37. Gombart, A.F., Shiohara, M., Kwok, S.H., Agematsu, K., Komiyama, A. and Koeffler, H.P. (2001) Neutrophil-specific granule deficiency: homozygous recessive inheritance of a frameshift mutation in the gene encoding transcription factor CCAAT/enhancer binding protein–epsilon. *Blood*, **97**, 2561–2567.
38. 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. Nat. Acad. Sci. U.S.A.*, **94**, 13187–13192.
39. Papaemmanuil, E., Hosking, F.J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., Kinsey, S.E., Lightfoot, T., Roman, E., Irving, J.A. *et al.* (2009) Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat. Genet.*, **41**, 1006–1010.
40. Trevino, L.R., Yang, W., French, D., Hunger, S.P., Carroll, W.L., Devidas, M., Willman, C., Neale, G., Downing, J., Raimondi, S.C. *et al.* (2009) Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat. Genet.*, **41**, 1001–1005.
41. Walsh, K.M., Chokkalingam, A.P., Hsu, L.I., Metayer, C., de Smith, A.J., Jacobs, D.I., Dahl, G.V., Loh, M.L., Smirnov, I.V., Bartley, K. *et al.* (2013) Associations between genome-wide

Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children. *Leukemia*, **27**, 2416–2419.
42. Walsh, K.M., de Smith, A.J., Welch, T.C., Smirnov, I., Cunningham, M.J., Ma, X., Chokkalingam, A.P., Dahl, G.V., Roberts, W., Barcellos, L.F. *et al.* (2014) Genomic ancestry and somatic alterations correlate with age at diagnosis in Hispanic children with B-cell acute lymphoblastic leukemia. *Am. J. Hematol.*, **89**, 721–725.
43. Weisberg, S.P., Smith-Raska, M.R., Esquilin, J.M., Zhang, J., Arenzana, T.L., Lau, C.M., Churchill, M., Pan, H., Klinakis, A., Dixon, J.E. *et al.* (2014) ZFX controls propagation and prevents differentiation of acute T-lymphoblastic and myeloid leukemia. *Cell Rep.*, **6**, 528–540.
44. Hunger, S.P. and Mullighan, C.G. (2015) Redefining ALL classification: toward detecting high-risk ALL and implementing precision medicine. *Blood*, **125**, 3977–3987.
45. Lee, Y.J., Jones, L.C., Timchenko, N.A., Perrotti, D., Tenen, D.G. and Kogan, S.C. (2006) CCAAT/enhancer binding proteins alpha and epsilon cooperate with all-trans retinoic acid in therapy but differ in their antileukemic activities. *Blood*, **108**, 2416–2419.
46. Chakraborty, P.K., Schmitz-Abe, K., Kennedy, E.K., Mamady, H., Naas, T., Durie, D., Campagna, D.R., Lau, A., Sendamarai, A.K., Wiseman, D.H. *et al.* (2014) Mutations in TRNT1 cause congenital sideroblastic anemia with immunodeficiency, fevers, and developmental delay (SIFD). *Blood*, **124**, 2867–2871.
47. Ito, T., Ando, H., Suzuki, T., Ogura, T., Hotta, K., Imamura, Y., Yamaguchi, Y. and Handa, H. (2010) Identification of a primary target of thalidomide teratogenicity. *Science*, **327**, 1345–1350.
48. Giagounidis, A.A. (2012) Lenalidomide for del(5q) and non-del(5q) myelodysplastic syndromes. *Semin. Hematol.*, **49**, 312–322.
49. Kronke, J., Udeshi, N.D., Narla, A., Grauman, P., Hurst, S.N., McConkey, M., Svinkina, T., Heckl, D., Comer, E., Li, X. *et al.* (2014) Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science*, **343**, 301–305.
50. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A. *et al.* (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, **491**, 119–124.
51. Moffatt, M.F., Gut, I.G., Demenais, F., Strachan, D.P., Bouzigon, E., Heath, S., von Mutius, E., Farrall, M., Lathrop, M., Cookson, W.O., Gabriel Consortium. (2010) A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med.*, **363**, 1211–1221.
52. Feng, B.J., Sun, L.D., Soltani-Arabshahi, R., Bowcock, A.M., Nair, R.P., Stuart, P., Elder, J.T., Schrodi, S.J., Begovich, A.B., Abecasis, G.R. *et al.* (2009) Multiple loci within the major histocompatibility complex confer risk of psoriasis. *PLoS Genet.*, **5**, e1000606.
53. 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.
54. Shimizu, R. and Yamamoto, M. (2016) GATA-related hematologic disorders. *Exp. Hematol.*, **44**, 696–705.
55. Nadif, R., Zerimech, F., Bouzigon, E. and Matran, R. (2013) The role of eosinophils and basophils in allergic diseases considering genetic findings. *Curr. Opin. Allergy Clin. Immunol.*, **13**, 507–513.
56. Rosser, F.J., Forno, E., Cooper, P.J. and Celedon, J.C. (2014) Asthma in Hispanics. An 8-year update. *Am. J. Respir. Crit. Care Med.*, **189**, 1316–1327.
57. Lavange, L.M., Kalsbeek, W.D., Sorlie, P.D., Aviles-Santa, L.M., Kaplan, R.C., Barnhart, J., Liu, K., Giachello, A., Lee, D.J., Ryan, J. *et al.* (2010) Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann. Epidemiol.*, **20**, 642–649.
58. Sorlie, P.D., Aviles-Santa, L.M., Wassertheil-Smoller, S., Kaplan, R.C., Daviglus, M.L., Giachello, A.L., Schneiderman, N., Raij, L., Talavera, G., Allison, M. *et al.* (2010) Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann. Epidemiol.*, **20**, 629–641.
59. Laurie, C.C., Doheny, K.F., Mirel, D.B., Pugh, E.W., Bierut, L.J., Bhangale, T., Boehm, F., Caporaso, N.E., Cornelis, M.C., Edenberg, H.J. *et al.* (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.*, **34**, 591–602.
60. Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. and McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
61. Conomos, M.P., Laurie, C.A., Stilp, A.M., Gogarten, S.M., McHugh, C.P., Nelson, S.C., Sofer, T., Fernandez-Rhodes, L., Justice, A.E., Graff, M. *et al.* (2016) Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *Am. J. Hum. Genet.*, **98**, 165–184.
62. Howie, B., Marchini, J. and Stephens, M. (2011) Genotype imputation with thousands of genomes. *G3 (Bethesda)*, **1**, 457–470.
63. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
64. 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. *Genetic Epidemiol.*, **34**, 816–834.
65. Chen, H., Wang, C., Conomos, M.P., Stilp, A.M., Li, Z., Sofer, T., Szpiro, A.A., Chen, W., Brehm, J.M., Celedon, J.C. *et al.* (2016) Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am. J. Hum. Genet.*, **98**, 653–666.
66. Reiner, A.P., Beleza, S., Franceschini, N., Auer, P.L., Robinson, J.G., Kooperberg, C., Peters, U. and Tang, H. (2012) Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. *Am. J. Hum. Genet.*, **91**, 502–512.
67. Chen, W., Brehm, J.M., Manichaikul, A., Cho, M.H., Boutaoui, N., Yan, Q., Burkart, K.M., Enright, P.L., Rotter, J.I., Petersen, H. *et al.* (2015) A genome-wide association study of chronic obstructive pulmonary disease in Hispanics. *Ann. Am. Thorac. Soc.*, **12**, 340–348.
68. Manichaikul, A., Palmas, W., Rodriguez, C.J., Peralta, C.A., Divers, J., Guo, X., Chen, W.M., Wong, Q., Williams, K., Kerr, K.F. *et al.* (2012) Population structure of Hispanics in the United States: the multi-ethnic study of atherosclerosis. *PLoS Genet.*, **8**, e1002640.
69. Tayo, B.O., Teil, M., Tong, L., Qin, H., Khitrov, G., Zhang, W., Song, Q., Gottesman, O., Zhu, X., Pereira, A.C. *et al.* (2011) Genetic background of patients from a university medical

center in Manhattan: implications for personalized medicine. *PLoS One*, **6**, e19166.
70. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
71. Sofer, T., Heller, R., Bogomolov, M., Avery, C.L., Graff, M., North, K.E., Reiner, A.P., Thornton, T.A., Rice, K., Benjamini, Y. *et al*. (2016) A powerful statistical framework for generalization testing in GWAS, with application to the HCHS/SOL. *Genet. Epidemiol*., 00, 1-11. Epub ahead of print.
72. Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
73. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R. *et al*. (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*., **44**, D110–D115.