# Genome-wide Association Study of Platelet Count Identifies Ancestry-Specific Loci in Hispanic/Latino Americans

Ursula M. Schick,[1,2,3,16] Deepti Jain,[4,16] Chani J. Hodonsky,[5,16] Jean V. Morrison,[4] James P. Davis,[6] Lisa Brown,[4] Tamar Sofer,[4] Matthew P. Conomos,[4] Claudia Schurmann,[2,3] Caitlin P. McHugh,[4] Sarah C. Nelson,[4] Swarooparani Vadlamudi,[6] Adrienne Stilp,[4] Anna Plantinga,[4] Leslie Baier,[7] Stephanie A. Bien,[1] Stephanie M. Gogarten,[4] Cecelia A. Laurie,[4] Kent D. Taylor,[8,9] Yongmei Liu,[10] Paul L. Auer,[11] Nora Franceschini,[5] Adam Szpiro,[4] Ken Rice,[4] Kathleen F. Kerr,[4] Jerome I. Rotter,[8] Robert L. Hanson,[7] George Papanicolaou,[12] Stephen S. Rich,[13,14] Ruth J.F. Loos,[2,3,15] Brian L. Browning,[4] Sharon R. Browning,[4] Bruce S. Weir,[4] Cathy C. Laurie,[4] Karen L. Mohlke,[6] Kari E. North,[5,16] Timothy A. Thornton,[4,16] and Alex P. Reiner[1,16,*]

Platelets play an essential role in hemostasis and thrombosis. We performed a genome-wide association study of platelet count in 12,491 participants of the Hispanic Community Health Study/Study of Latinos by using a mixed-model method that accounts for admixture and family relationships. We discovered and replicated associations with five genes (*ACTN1*, *ETV7*, *GABBR1-MOG*, *MEF2C*, and *ZBTB9-BAK1*). Our strongest association was with Amerindian-specific variant rs117672662 (p value = $1.16 \times 10^{-28}$) in *ACTN1*, a gene implicated in congenital macrothrombocytopenia. rs117672662 exhibited allelic differences in transcriptional activity and protein binding in hematopoietic cells. Our results underscore the value of diverse populations to extend insights into the allelic architecture of complex traits.

## Introduction

Platelets are small, anucleate cells derived from megakaryocyte cytoplasm in the bone marrow. Platelet production results from a series of tightly regulated processes that require lineage commitment of hematopoietic stem cells and leads to the proliferation, terminal differentiation, and maturation of megakaryocytic progenitors. Studying the genetic underpinnings of platelet count (PLT) can provide important insight into molecular mechanisms and pathways involved in both normal and abnormal megakaryopoiesis, which could ultimately have clinical implications for the treatment of bleeding or thrombosis in individuals with a low (thrombocytopenia) or high (thrombocytosis) PLT[1–3] or for the relationship between PLT and cardiovascular or autoimmune disorders.[4,5]

Circulating PLT in humans normally ranges between 150,000/μl and 400,000/μl. PLT differs by ethnicity, and these ethnic differences do not appear to be explained by environmental factors.[6,7] Family-based studies have estimated that a large component of the variability of PLT is explained by genetic factors ($h^2 > 0.50$).[8–10] To date, approximately 60 PLT-associated genetic variants have been identified through genome-wide association studies (GWASs) in populations of European, Asian, and African descent.[2,5,11–14]

Hispanic and/or Latino (Hispanic/Latino) individuals are a highly heterogeneous population with recent admixture among indigenous Amerindian (primarily of South and Central America, Mexico, and the Caribbean islands, hereafter referred to as "Amerindian"), European, and West African ancestral populations. Genetic factors contributing to PLT among Hispanic/Latino populations have not previously been characterized. Notably, certain Mendelian platelet disorders are more common among Hispanic/Latino individuals,[15] suggesting the possibility of population-specific genetic contributions to platelet-related phenotypes. To further characterize the role of genetic factors contributing to PLT in Hispanic/Latino populations, we performed a GWAS in 12,491 participants

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98195, USA; [2]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [3]Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [4]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [5]Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27514, USA; [6]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA; [7]Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Disease, NIH, 445 North 5th Street, Phoenix, AZ 85004, USA; [8]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [9]Department of Pediatrics, Los Angeles Biomedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [10]School of Medicine, Wake Forest University, Winston-Salem, NC 27157, USA; [11]Joseph J. Zilber School of Public Health, University of Wisconsin Milwaukee, Milwaukee, WI 53201, USA; [12]Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA; [13]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA; [14]Division of Endocrinology, Department of Medicine, University of Virginia, Charlottesville, VA 22908, USA; [15]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[16]These authors contributed equally to this work
*Correspondence: apreiner@uw.edu
http://dx.doi.org/10.1016/j.ajhg.2015.12.003. 

in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). We sought to identify genetic loci associated with PLT and assess generalization of known loci from other populations to this diverse sample of Hispanic/Latino individuals.

## Material 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 and selected from households in predefined census-block groups across four US field centers (in Chicago, Miami, the Bronx, and San Diego). The census-block groups were chosen to provide diversity among cohort participants with regard to socioeconomic status and national origin or background. The HCHS/SOL cohort includes participants who self-identified as having a Hispanic/Latino background; the largest groups are 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.[16] The HCHS/SOL baseline clinical examination[17] occurred between 2008 and 2011 and included comprehensive biological, behavioral, and sociodemographic assessments. This study was approved by the institutional review boards at each field center, where all subjects gave written informed consent.

### Measurement of PLT and Exclusion Criteria in HCHS/SOL

PLT was measured in EDTA whole blood with 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 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) ancestry-informative markers, known GWAS hits and drug absorption, distribution, metabolism, and excretion (ADME) markers, and additional custom content including ~150,000 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.[18]

We applied standardized quality-assurance and quality-control (QA/QC) methods[19] to generate recommended SNP- and sample-level quality filters. In brief, samples were checked for annotated or genetic sex, gross chromosomal anomalies,[20] relatedness[21] and population structure,[22] missing call rates, batch effects, and duplicate-sample discordance. At the SNP level, checks were performed for Hardy-Weinberg equilibrium, minor allele frequency (MAF), duplicate-probe discordance, Mendelian errors, and missing call rate. These QA/QC procedures yielded a total of 12,803 unique study participants for imputation and downstream association analyses. Of these, 12,491 met specific inclusion criteria related to the study of PLT. A total of 2,232,944 SNPs passed filters for both quality and informativeness (polymorphic and unduplicated) and became candidates for imputation and association testing.

### Imputation in HCHS/SOL

Genome-wide imputation was carried out with the full, cosmopolitan 1000 Genomes Project phase 1 reference panel (n = 1,092).[23] The HCHS/SOL samples were imputed together with genotyped SNPs passing the quality filter and representing unique genomic positions on the autosomes and non-pseudoautosomal portion of the X chromosome. Genotypes were first pre-phased with SHAPEIT2 (v.2.r644) and then imputed with IMPUTE2 (v.2.3.0).[24,25] Only variants with at least two copies of the minor allele present in any of the four 1000 Genomes continental panels were imputed. In addition to calculating the quality metrics output by IMPUTE2, we also calculated "oevar" (the ratio of the observed variance of imputed dosages to the expected binomial variance) by using the MaCH imputation software.[26] We assessed overall imputation quality both by looking at the distribution of imputed quality metrics across the MAF spectrum and by examining results from the IMPUTE2 internal masking experiments. We performed downstream association analyses only on observed variants passing quality filters and all imputed variants (a total of 27,887,661 variants), but we filtered the results on the basis of imputation quality (oevar > 0.3) and MAF > 1%.

### Linear Mixed-Effect Model for Association Testing in HCHS/SOL

We analyzed PLT by using linear mixed-effect models (LMMs) to account for the correlations due to genetic relatedness (kinship), shared household, and block group between individuals. The LMM used three independent random effects to model these three sources of dispersion:

$$y_i = x_i^T \alpha + g_{ij}\beta_j + b_{ik} + b_{ih} + b_{ib} + \varepsilon_i,$$

where $y_i$ is the square-root-transformed platelet value for individual $i$, $x_i$ is a vector of covariate values, $\alpha$ is the corresponding regression parameters, $g_{ij}$ is the $j^{\text{th}}$ SNP count, where $\beta_j$ is its estimated effect, and $b_{ik}$, $b_{ih}$, and $b_{ib}$ are the random effects corresponding to kinship, household, and block group, respectively (independent of each other and the error term $\varepsilon_i$), of person $i$. Within the LMM framework, $b_{ih}$ ($b_{ib}$) is the same among individuals who live in the same household (block group); for two individuals $i$ and $l$, the correlation between $b_{ik}$ and $b_{lk}$ is given by their estimated kinship coefficient.[27] The covariates included sex, age, principle components (PCs), recruitment center, smoking, log of sampling weight, and genetic-analysis group (a six-level categorical variable derived from self-identified background). With square-root-transformed PLT, the null-model residuals, given by $e_i = y_i - x_i^T\alpha - g_{ij}\beta_j$, were approximately normally distributed and thus compatible with modeling assumptions. We evaluated the goodness of fit of the LMM by using quantile-quantile plots comparing residuals and estimated random effects to a normal distribution and scatter plots describing the relationship between model residuals and covariates.

### Ancestry and Relatedness Adjustment in HCHS/SOL

We adjusted analyses for five PCs to prevent spurious association due to population stratification. Analyses accounted for familial

relatedness (kinship) by using a random effect with correlation structure specified by pairwise kinship coefficients for preventing inflation of test statistics. The PCs and kinship coefficients were estimated simultaneously with an iterative procedure, alternating between PC-AiR[22] (which provides PCs robust to familial relatedness) and PC-Relate[21] (which estimates kinship coefficients robust to population structure, admixture, and non-random mating). PC-AiR uses relatedness estimates to identify a mutually unrelated subset of individuals representative of the ancestral diversity of the entire sample, performs PCA on this unrelated subset, and predicts PC values for the remaining individuals. PC-Relate uses PCs to account for genetic similarity due to shared ancestry and provide accurate estimates of kinship coefficients due to familial relatedness. We performed three iterations, each of which used ~150,000 linkage-disequilibrium (LD)-pruned SNPs.[28]

### SNP-Based Heritability Estimation in HCHS/SOL

Genetic (kinship) and shared environmental (household) effects were estimated from a variance-component analysis that used all genotyped SNPs with MAF $> 1\%$ (~1.7 million) and a subset of 10,093 individuals estimated to be more distant than fourth-degree relatives (i.e., for whom all pairwise kinship coefficient estimates from PC-Relate were less than $2^{-11/2} \approx 0.022$). Including close relatives in the analysis can lead to inflated heritability estimates as a result of their increased phenotypic correlations due to other factors such as shared environmental effects.[29] However, the availability of current household membership data in HCHS/SOL made it possible that the variance-component model could at least partially account for shared environmental effects; therefore, the analysis was repeated with all 12,491 study individuals.

### Estimation of SNP Alleles and Allelic Frequencies among Ancestral Populations

We compared allele frequencies of PLT-associated index SNPs across ancestral Hispanic/Latino populations by using data from phase 3 of 1000 Genomes.[23] We used the R[31] exactci package to calculate exact p values and matching 95% confidence intervals (CIs) for each sub-population from the binomial distribution.[30] We also examined whether the derived alleles at our PLT-associated index SNPs were present in other ancestral human or Amerindian populations by using published whole-genome sequence data from Neandertal and Denisovan archaic human samples[32,33] and Papua New Guinea samples.[34]

### 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 3,454 from the Women's Health Initiative (WHI) SNP Health Association Resource (SHARe) project,[35] 782 from the Multi-Ethnic Study of Atherosclerosis (MESA) cohort,[36,37] and 2,854 from Mount Sinai BioMe Biobank.[38] WHI-SHARe and MESA participants were genotyped with the Affymetrix 6.0 chip, and imputation was performed with MaCH.[26] BioMe participants were genotyped with the Illumina HumanOmniExpressExome-8 v.1.0 chip, and imputation was performed with IMPUTE2[24,25] in 1000 Genomes phase 1 data (March 2012 v.3). Association testing for typed or imputed SNPs was performed by linear regression of square-root-transformed PLT adjusted for age, sex, and PCs.

### Meta-analysis and Replication Significance Criteria

Meta-analysis of results from the three replication cohorts for PLT and mean platelet volume (MPV) was performed with the inverse-variance-weighted method implemented in METAL.[39] To declare significance for replicated PLT loci, we used Bonferroni correction for the six variants carried forward for a significance threshold of p value $< 0.0083$.

### Admixture Mapping in HCHS/SOL

We implemented a conditional-random-field-based approach, RFMix,[40] to infer local ancestry at a set of 236,456 SNPs in common between the HCHS/SOL and reference-panel datasets. We used selected populations from HGDP,[41] HapMap 3,[42] and 1000 Genomes[23] phase 1 to use as a reference panel for detecting European, West African, and Amerindian ancestry. RFMix requires phased data with no missing genotype values. BEAGLE (v.4) was employed for phasing and imputation of sporadic missing genotypes in the HCHS/SOL and reference-panel datasets.[43] Admixture mapping is a powerful gene-mapping approach that relies on allele-frequency differences across ancestral populations and the existence of an association between the causal variant and phenotype to identify an association. Using the local-ancestry estimates, we performed a genome-wide admixture-mapping scan by using a LMM[43] with a joint test for all three ancestries (European, African, and Amerindian). As a secondary analysis, we performed admixture mapping to test Amerindian against any other ancestry because a priori we were interested in Amerindian ancestry within HCHS/SOL because it has not been well studied in previous admixture-mapping studies. Covariate and ancestry adjustment used in the analyses is described above. The recent history of admixture gives rise to long-range correlation in local-ancestry values across the genome, and thus the critical value for the genome-wide significance level of admixture mapping is substantially lower than that for the genotype test. On the basis of previous simulation results, a nominal p value of $5.7 \times 10^{-5}$ yielded a genome-wide type I error of 0.05.

### Generalization in HCHS/SOL

We performed generalization analysis for PLT-associated SNPs previously reported in GWASs of other populations, including those of European, African, and Japanese ancestry.[2,5,11–13] Because all discovery studies, except for that of Kamatani et al.,[11] used untransformed PLT as the outcome, we used association results with untransformed PLT for generalization. For SNPs reported in Kamatani et al.,[11] we followed their methodology and used the square root of PLT and reported effect sizes in SDs. We performed generalization testing by directional false-discovery rate (FDR) control for the generalization null hypotheses.[44] The generalization null hypothesis states that the effect does not exist in both the discovery study and 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. We used the number of SNPs tested in the discovery study and the p values for the set of tested SNPs from both the discovery study and HCHS/SOL, and we computed an r value for each of the SNPs to quantify the evidence for generalization. A SNP was generalized if the r value $< 0.05$.

### Functional Annotation of Discovery Loci

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

($r^2 \geq 0.5$, calculated in the HCHS/SOL discovery population) were positioned within predicted regulatory regions, namely enhancers and promoters. These regulatory regions were identified on the basis of the enrichment of various histone-modification and ChIP-seq (chromatin immunoprecipitation followed by sequencing) signals in megakaryocytes.[45] 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 that belonged to either promoter or enhancer categories and overlapped a DNase I hypersensitive site (a general biochemical feature of regulatory regions) in megakaryocytes were prioritized as putatively functional variants. We also reported overlap with ChIP-seq peaks of key megakaryocyte transcription factors and the nearest biologically plausible gene or genes.[46] Moreover, because related cell types can share similar regulatory regions, we additionally reported supplementary annotation by using data generated by ENCODE[47] on other myeloid lineage cells, including primary erythroblasts, erythroleukemia K562 cells, peripheral-blood-derived erythroblasts (PBDEs), myeloid leukemia (SKNO-1) cells, and human umbilical vein endothelial cells (HUVECs). To provide additional support, we also included overlap with transcription start sites and enhancers identified by an alternate approach in the Fantom5 project. [48] To identify the motifs disrupted by alleles, including *ACTN1* (MIM: 102575) variant rs117672662, we utilized HaploReg (v.2)[49] and the JASPAR motif database.[50]

We also included annotations from in silico prediction algorithms including RegulomeDB, the Combined Annotation Dependent Depletion (CADD) score[51] (a PHRED-like score indicating deleteriousness of variants and all other substitutions in the genome), GWAVA[52] (a score that classifies non-coding variation and uses ENCODE and Roadmap Epigenomic data to prioritize most likely functional variants), and deltaSVM[53] (a score that captures how much a variant alters the regulatory potential of the surrounding sequence, particularly in the context of a specific cell type).

### eQTL Analysis of American Indians

The eQTL analysis included 1,457 American Indian adults (minimum of 18 years of age) from the urban Phoenix extension of the Family Investigation of Nephropathy and Diabetes; they were examined after they had fasted for ≥8 hr.[54] Blood was collected into PAXgene Blood RNA Tubes (Becton Dickinson), and total RNA was isolated with PAXgene Blood miRNA Kits (QIAGEN). Amplification was performed with the Ambion MessageAmp II-Biotin Enhanced aRNA Amplification Kit (Life Technologies), and transcript levels were measured with the Illumina HumanHT-12 v.4 Expression Beadchip according to the manufacturer's protocol. GenomeStudio software was used for background normalization. Genotyping of rs117672662 was conducted according to the Assays-on-Demand method (Life Technologies). A normalizing transformation of transcription levels was employed in statistical analyses, and association of genotype was analyzed under an additive model with control for age, sex, tribal membership, and European admixture (estimated from 45 markers with large allele-frequency differences between Amerindians and Europeans[55]).

### Cell Culture

THP-1 (ATCC TIB-202) acute monocytic leukemia cells were cultured in RPMI-1640 (Mediatech) supplemented with 10% fetal bovine serum (FBS), and Kasumi-1 (ATCC CRL-2724) acute myeloblastic leukemia cells were cultured in RPMI-1640 supplemented with 20% FBS. The cell cultures were maintained at 37°C with 5% $CO_2$.

### Transcriptional Reporter Assays

A 186 bp region (chr14: 69,425,369–69,425,554 according to UCSC Genome Browser hg19) surrounding rs117672662 was amplified with primer pairs 5′-GGTACCGCAGGAAAACATC CACATGA-3′ and 5′-CTCGAGGGAAACAGTGTGGTCAGTCG-3′ (forward) and 5′-CTCGAGGCAGGAAAACATCCACATGA-3′ and 5′- GGTACCGGAAACAGTGTGGTCAGTCG-3′ (reverse) and cloned into the luciferase reporter vector pGL4.23 (Promega) in both orientations with respect to the minimal promoter. The rs117672662 C allele was created with the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Sanger sequencing was used to verify clones for fidelity and genotype. Four verified constructs for each allele in both orientations were transfected in duplicate into THP-1 and Kasumi-1 cells with *Renilla* control reporter vector (phRL-TK, Promega) with Lipofectamine 3000 (Life Technologies) and incubated for 48 hr. The cells were lysed with Passive Lysis Buffer (Promega), and luciferase activity was measured with the Dual-Luciferase Reporter Assay System (Promega) as previously described.[56]

### Electrophoretic Mobility Shift Assay

Electrophoretic mobility shift assay (EMSA) oligonucleotide probes were designed around the variant rs117672662 (5′-AGAATTAT[T/C]AGCAGAGG-3′) and end labeled with 5′ IRDye 700 (Integrated DNA Technologies). Nuclear protein was extracted with the NE-PER Extraction Kit (ThermoFisher Scientific), and the total nuclear extract was measured with the BCA Protein Assay (ThermoFisher Scientific). All protein-probe binding reactions were incubated for 30 min at room temperature and consisted of the following: 1× binding buffer (10 mM Tris, 50 mM KCl, and 1 mM DTT [pH 7.5]), 1 μg poly(dI-dC), 7 μg nuclear extract, and 200 fmol IRDye-labeled double-stranded oligonucleotide probe in a volume of 20 μl. The competition reactions contained 70-fold excess of unlabeled probe and were incubated with the nuclear extract for 15 min prior to the addition of the IRDye-labeled probe and incubation for another 30 min. In a test for an antibody supershift, 6 μg of antibody (HOXA5 sc-13199x or GATA1 sc-1234x, Santa Cruz Biotechnologies) was incubated with the nuclear extract for 35 min prior to incubation with the IRDye-labeled probe for 30 min. The probe-protein complexes were resolved with 6% DNA retardation electrophoresis gels (Life Technologies) and visualized with an Odyssey CLx Infrared Imaging System (LI-COR Biosciences).

## Results

The characteristics of the 12,491 Hispanic/Latino participants from the HCHS/SOL discovery sample are summarized in Table S1. SNP-based heritability was estimated from a variance-component analysis performed with a subset of 10,093 individuals excluding close familial relatives. Genetic (kinship) effects accounted for 29.4% (95% CI: 22.6%–36.1%) of the variation in PLT, whereas shared environmental (household) effects contributed

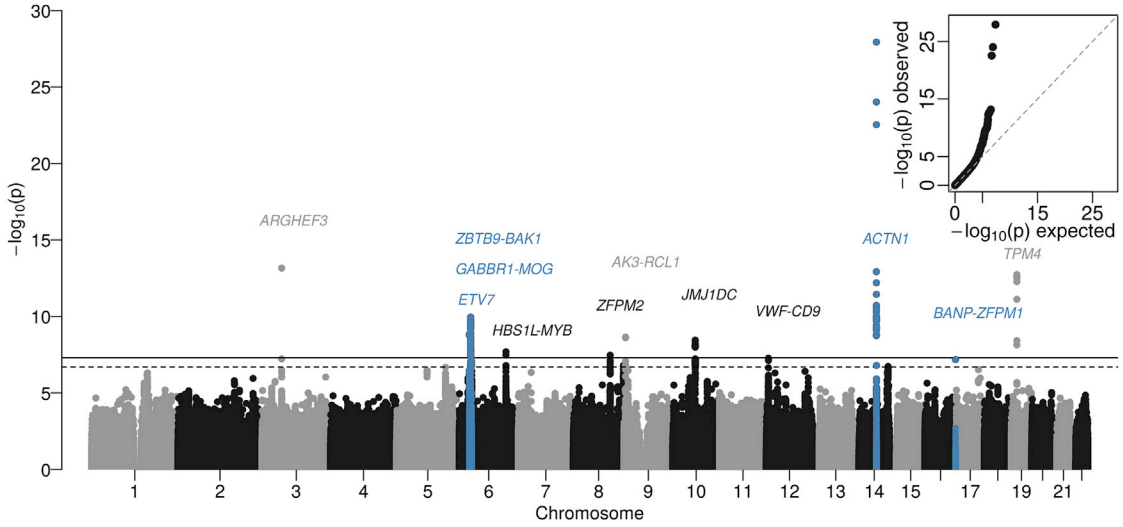



**Figure 1. Manhattan Plot of Discovery Results from HCHS/SOL**
The solid line indicates genome-wide significance (p value $< 5 \times 10^{-8}$), and a dashed line indicates suggestive significance (p value $< 1 \times 10^{-7}$). There is an inset quantile-quantile plot of discovery p values. Discovery 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).

little (2.6%; 95% CI: 0.0%–6.3%). When the analysis was repeated with all individuals and adjustment for household-membership data, genetic (kinship) effects accounted for 30.8% (95% CI: 25.7%–35.7%), and shared environmental (household) effects accounted for 5.0% (95% CI: 2.1%–7.7%) of PLT variation. Thus, the estimated genetic contribution increased only slightly (from 29.4% to 30.8%) when close relatives were included, and the estimated household contribution was non-zero, suggesting that household membership is a good proxy for the shared environmental effects contributing to PLT in this sample. These heritability estimates are less than what has been reported from family-based studies of PLT (approximately 50%–80%[8–10]), which is consistent with previous findings comparing family- and SNP-based heritability estimation.[57]

In the HCHS/SOL discovery sample, the genomic inflation factor was 1.046, indicating adequate control of population stratification. Nine loci met the standard significance criteria of p value $< 5 \times 10^{-8}$; three additional loci had p values between $5 \times 10^{-8}$ and $1 \times 10^{-7}$ (Table S2). Quantile-quantile and Manhattan plots are shown in Figure 1. For each of the discovery and previously reported PLT loci, we evaluated the extent of LD (calculated HCHS/SOL discovery population) between the index SNP and other nominally significant SNPs in the region (Figures S1A–S1L).

### Generalization of PLT Index SNPs from Other Populations to Hispanic/Latino Populations

Of the 12 genome-wide significant or suggestive loci in the HCHS/SOL discovery sample, seven correspond to index SNPs (or LD proxies with $r^2 \geq 0.5$ calculated in the HCHS/SOL population) previously identified in PLT GWASs of other ancestries (*ARHGEF3* [MIM: 612115] rs1354034, *TPM4* [MIM: 600317] rs73517714, *AK3* [MIM: 609290]-*RCL1* [MIM: 611405] rs409801, *JMJD1C* [MIM: 604503] rs10822155, *ZFPM2* [MIM: 603693] rs6993770, *HBS1L* [MIM: 612450]-*MYB* [MIM: 189990] rs6934903, and *CD9* [MIM: 143030]-*VWF* [MIM: 613160] rs11064074)[2,5,12–14] (Table S2). An eighth significant locus is located in close proximity to *BAK1* (MIM: 600516), a gene previously associated with PLT in GWASs from populations of European, African, and Asian descent.[2,5,11,12,14] However, the index SNP (rs62405954) in our Hispanic/Latino discovery sample at the *BAK1* locus is distinct from rs210134, the index SNP previously associated with PLT in a GWAS ($r^2 = 0.06$; rs62405954 p value = $1.10 \times 10^{-10}$, versus $4.6 \times 10^{-7}$ when adjusted for rs210134; Figure S2).

In order to more comprehensively assess whether previous GWAS PLT SNPs from populations of European, Asian, and African ancestry[2,5,11–14] generalize to HCHS/SOL Hispanic/Latino populations, we evaluated all index SNPs in the corresponding ancestral populations (Table S3) by using a directional FDR method[58] that rejects the null hypothesis of "no generalization" if there is enough evidence that a SNP is associated with PLT and directionally consistent between the original discovery GWAS and HCHS/SOL. Of the ten SNPs identified previously in populations of African ancestry,[12,13] seven of the SNPs showed evidence of generalization (r value $< 0.05$) in our Hispanic/Latino sample. Roughly half (27 of 49) of GWAS SNPs identified previously in populations of European ancestry[2,5,12,14] generalized to HCHS/SOL. All four SNPs identified previously in a population of Asian ancestry[11] generalized to our study. Considering the 55 independent SNPs previously associated with PLT in any population, we found evidence of generalization of 30

**Table 1. PLT-Associated Variants from the HCHS/SOL Discovery Analysis**

| Nearest Gene (Function) | Chromosomal Position (GRCh37/hg19) | rsID | Coded/Alternative Allele on Plus Strand | Discovery | | | | Replication | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Coded Allele Frequency | n | Beta (SE) | p Value | n | Beta (SE) | p Value |
| *ACTN1* (intronic) | chr14: 69,425,467 | rs117672662 | T/C | 0.94 | 12,491 | 0.604 (0.054) | $1.16 \times 10^{-28}$ | 7,121 | 0.685 (0.080) | $1.07 \times 10^{-17}$ |
| *ZBTB9-BAK1* (intergenic) | chr6: 33,524,820 | rs62405954 | T/C | 0.86 | 12,491 | −0.239 (0.037) | $1.10 \times 10^{-10}$ | 7,170 | −0.268 (0.066) | $4.93 \times 10^{-5}$ |
| *GABBR1-MOG* (intergenic) | chr6: 29,608,184 | rs75140056 | C/CAT | 0.39 | 12,491 | −0.151 (0.025) | $1.47 \times 10^{-9}$ | 7,163 | −0.236 (0.041) | $9.82 \times 10^{-9}$ |
| *BANP-ZFPM1* (intergenic) | chr16: 88,376,014 | rs80294974 | G/A | 0.98 | 12,491 | 0.555 (0.103) | $6.60 \times 10^{-8}$ | 7,060 | −0.082 (0.138) | 0.551 |
| *ETV7* (intronic) | chr6: 36,344,980 | rs9470264 | G/A | 0.80 | 12,491 | −0.181 (0.034) | $8.80 \times 10^{-8}$ | 6,998 | −0.145 (0.048) | 0.00263 |
| *MEF2C* (intronic) | chr5: 88,133,921 | rs144261491 | C/T | 0.97 | 12,491 | 0.426 (0.083) | $3.35 \times 10^{-7}$ | 3,982 | 0.624 (0.156) | $6.32 \times 10^{-5}$ |

(55%) SNPs in the HCHS/SOL population. We hypothesized that some of the SNPs did not generalize as a result of low power. To study this hypothesis, we (1) looked for directional consistency of the effect sizes across our study and previous studies for SNPs that failed to generalize and (2) generated a genetic score for each of the analysis participants by summing all trait-increasing alleles in the SNPs that did not generalize. Out of 25 SNPs that did not generalize, 24 had the same direction of effect in the discovery study and the HCHS/SOL (exact binomial test p value = $1.5 \times 10^{-6}$). We found a strong association between PLT and the genetic score that we constructed with the non-generalized SNPs (p value = $8.8 \times 10^{-10}$). Taken together, these tests provide evidence that, indeed, the majority of non-generalized SNPs are associated with PLT.

### Discovery of Ancestry-Specific PLT Association Signals in Hispanic/Latino Populations

Four of the 12 loci with significant or suggestive associations in the HCHS/SOL discovery sample (*ACTN1* rs117672662, *GABBR1* [MIM: 603540]-*MOG* [MIM: 159465] rs75140056, *ETV7* [MIM: 605255] rs9470264, and *BANP* [MIM: 611564]-*ZFPM1* [MIM: 601950] rs80294974) are located within or near genes or genomic regions not previously associated with PLT in GWASs (Table 1). By examining the 55 other genomic regions previously associated with PLT in populations of European, Asian, or African descent, in addition to *ZBTB9-BAK1* rs62405954 (described above), we identified one additional signal, *MEF2C* (MIM: 600662) rs144261491, distinct from the *MEF2C* European index SNP rs700585 from a GWAS ($r^2 = 0.009$; rs144261491 p value = $3.4 \times 10^{-7}$, versus $9.0 \times 10^{-7}$ with adjustment for rs700585; Figure S3).

Several of the PLT index SNPs discovered in HCHS/SOL had allele frequencies that differed considerably between continental populations represented in 1000 Genomes phase 3 samples[23] (Table 2). In particular, *ACTN1* rs117672662 and *MEF2C* rs14426149, associated with decreased PLT, were present at MAFs of ~7% and ~4%, respectively, in samples of Amerindian ancestry (Columbians, Mexicans, Peruvians, and Puerto Ricans) but were significantly less common in Asian, European, and African populations.[23] Analysis of published archaic genomes[32,33] and genome sequences from New Guinea Papuans[34] showed that none of the risk alleles at these six PLT loci appear to be derived from Neandertal, Denisovan, or Australo-Melanesian sequences.

Admixture mapping based on a joint test of all three local-ancestry estimates in HCHS/SOL confirmed the presence of genome-wide-significant peaks (p value < $5.7 \times 10^{-5}$) at the *BAK1* locus on chromosome 6 and a peak at chromosomal region q13.2, as well as a suggestive peak at the *ACTN1* locus on chromosome 14 (Figure S4A). Considering the secondary analysis comparing Amerindian ancestry to all other ancestries, we identified a

**Table 2. Allele Frequencies of PLT-Associated Variants by 1000 Genomes Continental Populations and Admixed American Sub-populations**

| Gene (Function) | Chromosomal Position (GRCh37/hg19) | rsID | Coded/Alternative Allele on Plus Strand | HCHS/SOL Coded Allele Frequency | Super-population Coded Allele Frequency | | | | AMR Sub-population Coded Allele Frequency | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | AFR | EAS | EUR | SAS | AMR | CLM | MXL | PEL | PUR |
| *ACTN1* (intronic)[a] | chr14: 69,425,467 | rs117672662 | T/C | 0.94 | 1.000 | 1.00 | 0.999 | 0.999 | 0.927 | 0.926 | 0.938 | 0.859 | 0.976 |
| *ZBTB9-BAK1* (intergenic) | chr6: 33,524,820 | rs62405954 | T/C | 0.86 | 0.999 | 0.994 | 0.919 | 0.964 | 0.831 | 0.824 | 0.781 | 0.741 | 0.942 |
| *GABBR1-MOG* (intergenic) | chr6: 29,608,184 | rs75140056 | C/CAT | 0.39 | 0.384 | 0.213 | 0.417 | 0.304 | 0.365 | 0.452 | 0.391 | 0.206 | 0.399 |
| *BANP-ZFPM1* (intergenic) | chr16: 88,376,014 | rs80294974 | G/A | 0.98 | 0.998 | 1.000 | 0.963 | 0.993 | 0.991 | 0.995 | 0.992 | 0.994 | 0.986 |
| *ETV7* (intronic) | chr6: 36,344,980 | rs9470264 | G/A | 0.80 | 0.825 | 0.877 | 0.992 | 0.995 | 0.765 | 0.809 | 0.656 | 0.612 | 0.918 |
| *MEF2C* (intronic)[a] | chr5: 88,133,921 | rs144261491 | C/T | 0.97 | 1.000 | 1.000 | 1.000 | 1.000 | 0.963 | 0.989 | 0.953 | 0.906 | 0.990 |

Abbreviations are as follows: AFR, African; EAS, East Asian; EUR, European; SAS, South Asian; AMR, admixed American; CLM, Colombian from Medellin, Colombia; MXL, Mexican ancestry from Los Angeles, USA; PEL, Peruvians from Lima, Peru; and PUR, Puerto Ricans from Puerto Rico.
[a]For *ACTN1* rs117672662 and *MEF2C* rs144261491, the allele frequencies differ significantly between AMR and AFR, EAS, EUR, and SAS populations.

significant peak at *ACTN1* (Figure S4B). Further, there was highly significant concordance between the number of Amerindian ancestral alleles at the *ACTN1* locus and the rs117672662 genotype (p value $< 2.20 \times 10^{-16}$; Table S4).

## Replication of PLT Loci in Independent Hispanic/Latino Samples

Replication of PLT association findings discovered in HCHS/SOL was carried out in an independent sample of up to 7,170 Hispanic/Latino Americans derived from three multi-ethnic US-based cohorts (WHI [n = 3,534], BioMe Biobank [n = 2,854], and MESA [n = 782]), whose characteristics are described in Table S1. We carried forward six SNPs from the discovery stage for replication: *ACTN1* (rs117672662), *GABBR1-MOG* (rs75140056), *ETV7* (rs9470264), *BANP-ZFPM1* (rs80294974), *ZBTB9-BAK1* (rs62405954), and *MEF2C* (rs144261491). Of these, five SNPs (all except *BANP-ZFPM1* rs80294974) met our prespecified criteria for replication (p value $< 0.05/6 = 0.008$; Table 1). *BANP-ZFPM1* rs80294974 had a lower frequency (MAF = 2%), so a failure to replicate might be related to limited power to detect this association in the smaller Hispanic/Latino replication sample. The discovery variants and the discovery and generalized variants from HCHS/SOL explain 1.79% and 6.14% of the total variance of PLT, respectively.

In addition to PLT, MPV measurements were available in a subset of 4,041 Hispanic/Latino WHI and BioMe participants in our replication dataset. At four out of five of our replicated loci, the allele associated with lower PLT was also associated with higher MPV (*ACTN1* rs117672662 p value = $3.90 \times 10^{-18}$, *GABBR1-MOG* rs75140056 p value = 0.003, *ETV7* rs9470264 p value = $1.90 \times 10^{-15}$, and *BAK1* rs62405954 p value = 0.03; Table S5).

## Functional Annotation and Characterization of the Discovery PLT Loci

At each of the five replicated loci (*ACTN1*, *GABBR1-MOG*, *ETV7*, *ZBTB9-BAK1*, and *MEF2C*), we defined the association interval as containing all genotyped or imputed variants (SNPs and indels) in LD ($r^2 > 0.5$) with the index variant. Within each interval, we (1) identified genes and non-coding RNAs and their tissue expression patterns, (2) predicted SNP functionality by using genome-wide epigenomic datasets from megakaryocytes and other blood cell types from the BLUEPRINT,[59] ENCODE,[47] and FANTOM5[60] projects (Table S6), and (3) assessed SNP associations with gene expression (eQTL) in whole blood[61] (Table S7).

We identified one or more SNPs that overlapped putative megakaryocyte enhancers or promoters and were in LD ($r^2 \geq 0.5$) with the index SNP at each of the five replicated PLT loci (Table S8). Of particular interest, *ACTN1* index SNP rs117672662 lies within a megakaryocyte-specific putative enhancer located within *ACTN1* intron 1 (Figure 2). The genomic element harboring rs117672662 overlaps ChIP-seq peaks of key megakaryocyte regulators, including

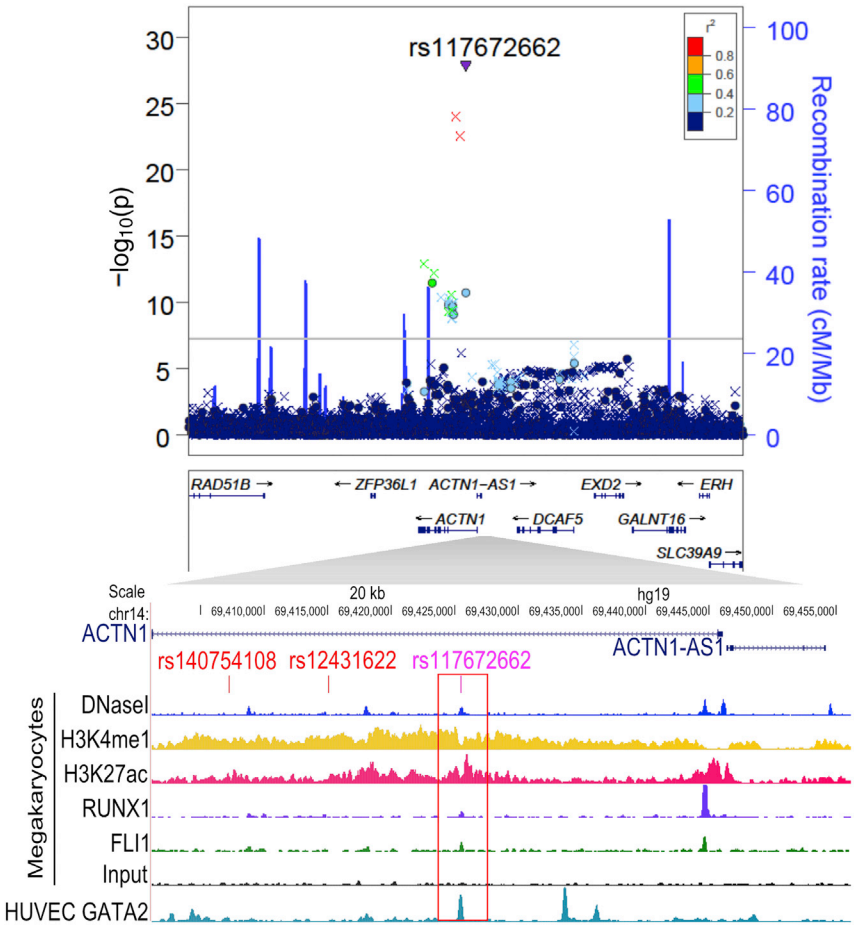



**Figure 2. Regional Plot of the *ACTN1* Locus**

The top panel contains a LocusZoom plot of the *ACTN1* locus centered on our top Amerindian-specific variant, rs117672662. The LD estimates derived from the HCHS/SOL study samples with respect to rs117672662 and the other variants in the window are color coded according to the scale indicated in the top panel. The imputed SNP, rs117672662, is denoted by a filled triangle, other imputed variants are denoted by an x, and genotyped variants are denoted by a filled circle. Recombination hotspots from HapMap are indicated by the vertical blue lines (see Web Resources). The horizontal line indicates the significance threshold p value $\leq 5 \times 10^{-8}$. The bottom panel is a UCSC Genome Browser screenshot zoomed in to show rs117672662 and its LD proxies ($r^2 > 0.8$). These variants are aligned with selected signal tracks of megakaryocytes, including DNaseI hypersensitivity, ChIP-seq for enhancer histone modifications H3K4me1 and H3K27ac, ChIP-seq for megakaryocyte transcription factors RUNX1 and Fli1, and the input (no antibody) track. A signal track of ChIP-seq for GATA2, another important megakaryocyte transcription factor in human umbilical vein endothelial cells (HUVECs), is also displayed. The red box is centered on the putative functional SNP rs117672662.

ERG, FLI1, and RUNX1 in SKNO-1 (acute myeloid leukemia) cells, further supporting its role as an enhancer. Furthermore, rs117672662 was predicted to have high regulatory potential according to the cell-type-specific regulatory-motif detection algorithm deltaSVM[53] trained on the myelogenous leukemia cell line K562 (score = 9.8; Table S9). The regulatory deltaSVM score for rs117672662 is in the same range as previous predictions for known functional SNPs.[53]

We performed de novo genotyping of the *ACTN1* rs117672662 variant in a sample of 1,457 American Indians who were from urban Phoenix and had previously undergone whole-blood transcriptomic analysis. The eQTL analysis did not reveal any significant eQTLs in the region for *ACTN1* (C allele: beta [SE] = −0.05 [0.085]; p value = 0.56) or for any other genes in the association interval (Table S10).

### Allelic Differences in Enhancer and Protein Binding Activity of *ACTN1* rs117672662

To further assess the regulatory properties of *ACTN1* variant rs117672662, we performed transcriptional reporter assays in THP-1 monocytic leukemia and Kasumi-1 myeloid leukemia cells. In both cell types, the rs117672662 minor C allele showed higher transcriptional activity than the rs117672662 T allele (Figure S5).

We used THP-1 cells to perform EMSAs and observed that the T allele showed stronger protein binding than the C allele (Figure 3). Including 70-fold excess T allele probe decreased the intensity of the protein-probe band more than including excess C allele probe, supporting allelic differences in specificity of the protein-probe binding (Figure 3 and Figure S6). To characterize the transcription factors binding to the rs1176762662 site, we conducted super-shift assays with antibodies of transcription factors whose DNA binding motifs had been predicted by Haploreg[49] to be altered by the variant. An overlapping HOXA5 motif matched better to the T allele, whereas GATA1 and GATA2 motifs matched better to the C allele. With THP-1 cell nuclear lysate, inclusion of HOXA5 antibodies shifted the protein complex bound to the T allele (Figure 3). The addition of GATA1 antibodies showed a partial shift, suggesting that a protein-probe complex might also include GATA1. Taken together, the transcriptional activity and gel-shift assays suggest that a protein complex binding to the T allele contains HOXA5 and might act as a transcriptional repressor of the target gene(s).

## Discussion

We report the results from a large GWAS of PLT in Hispanic/Latino populations. We identified and replicated

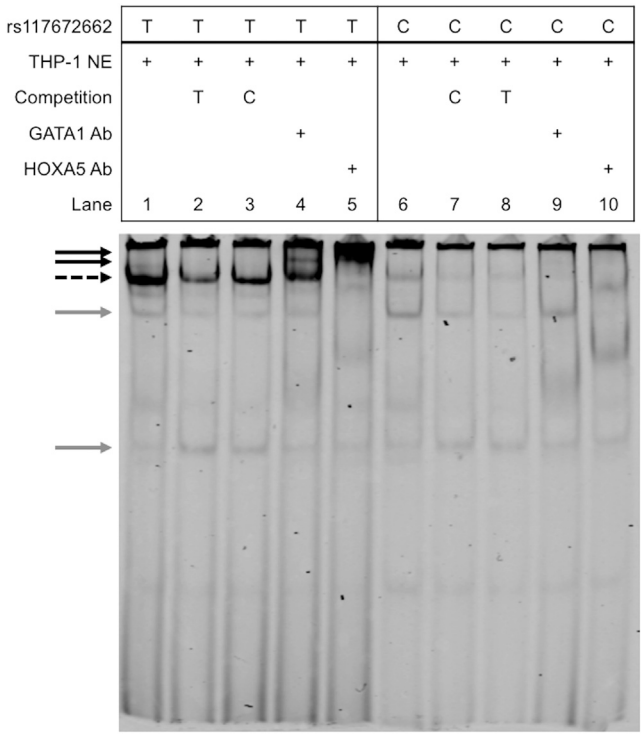



**Figure 3. Allelic Differences in Protein Binding at rs117672662**
EMSA using oligonucleotide probes containing different alleles at rs117672662 (T allele [lanes 1–5] and C allele [lanes 6–10]). Nuclear extracts from human monocyte THP-1 cells were incubated with IRDye-labeled double-stranded oligonucleotide probe alone (lanes 1 and 6) or with 70-fold excess of unlabeled probe (lanes 2, 3, 7, and 8), GATA1 antibodies (lanes 4 and 9), or HOXA5 antibodies (lanes 5 and 10). The dotted black arrow indicates probe-protein complexes, the solid black arrows indicate probe-protein-antibody complexes, and the gray arrows indicate non-specific probe-protein binding complexes. Further support of the allelic differences is provided in Figure S6.

associations with three loci including noncoding SNPs in or near *ACTN1*, *ETV7*, and *GABBR1-MOG* and two population-specific variants at previously identified PLT GWAS loci *ZBTB9-BAK1* and *MEFC2*. The *ACTN1* and *ZBTB9-BAK1* association signals were also detected in a genome-wide scan for local-ancestry admixture. Overall, four of the five PLT association signals (*ACTN1*, *ETV7*, *MEF2C*, and *ZBTB9-BAK1*) were highly differentiated across populations of European, West African, and Amerindian ancestry. The *ACTN1* and *MEF2C* alleles were found only on an Amerindian ancestral background and therefore could realistically only have been discovered through studies involving Hispanic/Latino or Amerindian populations. We have also demonstrated that approximately 50% of PLT-associated alleles previously identified in European, African American, and Japanese populations generalized to HCHS/SOL populations, suggesting that many of the same regions of the genome are involved in regulation of PLT across global populations.

The identification of a common Amerindian ancestral variant located in a putative enhancer region within intron 1 of *ACTN1* adds to previous studies reporting the *ACTN1* locus as a source of PLT phenotypic variation. The most likely targets of the rs117672662 variant are either *ACTN1* itself or *ACTN1-AS1*, which is a long non-coding RNA just upstream of *ACTN1* and can potentially regulate *ACTN1* transcript levels. Missense mutations in *ACTN1* have recently been identified in congenital macrothrombocytopenia pedigrees with mild-to-moderate thrombocytopenia, increased MPV, and minimal bleeding manifestations.[62,63] Two subsequent genetic studies (one from Italy and one from the US and Europe) of previously uncharacterized inherited platelet disorders found *ACTN1* missense mutations in ~4%–5% of affected individuals.[64,65] Functional characterization of missense variants in *ACTN1* (c.94C>A [p.Gln32Lys] and c.313G>A [p.Val105Ile]) suggest that the variants disrupt the actin cytoskeleton structure and impair megakaryocyte pro-platelet production.[63] Although these studies highlight the implications of *ACTN1* loss-of-function coding mutations in PLT regulation, transcriptional regulation of *ACTN1* is also important for normal megakaryopoiesis.[66,67] In addition to regulating PLT, the actin cytoskeleton is involved in determining platelet size during the final stages of pro-platelet formation from megakaryocytes.[68]

Consistent with the relationship between *ACTN1* mutations and familial macrothrombocytopenia, the minor allele (rs117672662 C allele) of the Amerindian *ACTN1* non-coding variant was associated with lower PLT and higher MPV. The rs117672662 C allele displayed increased enhancer activity, whereas the rs117672662 T allele demonstrated a super-shift that could be mediated by HOXA5 and a partial super-shift mediated by GATA1. GATA1 also appeared to bind to the rs117672662 T allele probe. HOXA5 and GATA1 have each been shown to play a role in erythrocyte and megakaryocyte development,[69–72] and both transcription factors are dysregulated in hematopoietic stem cells showing erythroid and megakaryocyte differentiation blockage mediated by *HOXA10* overexpression.[70] Furthermore, HOXA5 might act as a transcriptional repressor for several genes involved in actin remodeling.[73] Despite the evidence for allelic difference in expression and binding of HOXA5 and GATA1, further experiments are needed for elucidating the precise regulatory molecular mechanism by which the *ACTN1* rs117672662 C allele alters platelet production and/or PLT.[74,75]

Two additional PLT loci and one independent signal in a known locus were identified on chromosome 6 in gene-rich, extended LD regions located about 3.3 Mb centromeric (*ETV7*), 84 kb telomeric (*GABBR1-MOG*), and 25 Mb centromeric (*ZBTB9-BAK1*) to the major histocompatibility complex (MHC) region. The *ETV7* index SNP rs9470264 is located within an intron of *ETV7*, which encodes the Ets family transcription factor TEL-2. *ETV7* (or *TEL2*) is primarily expressed in hematopoietic tissues, including bone marrow and spleen.[76,77] Further, *ETV7* is expressed in leukemic cells and appears to be more highly expressed in a subset of leukemic samples,[78] pointing to possible roles for this transcription factor in both normal

hematopoiesis and oncogenesis. Mutations in a related Ets-encoding gene, *ETV6* (MIM: 600618), encoding ubiquitously expressed TEL-1, were recently identified in pedigrees affected by congenital thrombocytopenia,[79–81] and chromosomal translocations involving *ETV6* (e.g., *ETV6-PDGFRB* [MIM: 173410]) are common in hematologic malignancies.[78] Despite strong biological plausibility supporting *ETV7* in hematopoiesis, in silico functional annotation of the index SNP and LD proxies prioritized a promoter polymorphism in neighboring *KCTD20* (MIM: 615932) as the putative functional SNP. Whole-blood eQTL analysis of the *ETV7* index SNPs and LD proxies indicated significant differences in *KCTD20*, but not *ETV7*, expression (see Table S7), further supporting a role for *KCTD20* in regulating PLT at this locus.[61]

The index variant rs75140056 is located in an intergenic region between *MOG* and *GABBR1* (Figure S1C). Surveying variation in high LD with the index SNP, we prioritized rs29269 as the strongest functional candidate polymorphism on the basis of its position in a putative active *GABBR1* promoter in megakaryocytes (Table S8). *GABBR1* is a member of the gamma-aminobutyric acid family of inhibitory neurotransmitters and is most notably known for its role in the mammalian CNS. However, a recent study reported differential regulation of *GABBR1* in bone-marrow- and fetal-liver-derived megakaryocytes from wild-type mice, thereby suggesting a potential role of GABBR1 in developmental-stage-specific regulation of megakaryopoiesis.[82] In addition, several proxy SNPs, including putative functional SNP rs29269, are eQTLs (Table S7) for MHC class I genes (e.g., *HLA-F* [MIM: 143110] and *HLA-G* [MIM: 142871]). These observations suggest that variation at this locus might regulate one or more genes.

The index SNP in our Hispanic/Latino discovery sample at the *ZBTB9-BAK1* locus (rs62405954) is distinct from rs210134, which was previously associated with PLT in European GWASs. rs62405954 is in LD with rs1002011, which overlaps a DNase hypersensitive site in megakaryocytes and lies within a putative enhancer overlapping the 5′ UTR of *VPS52* (MIM: 603443), a gene that encodes an intracellular protein involved in endocytic recycling and is highly expressed in hematopoietic cells of erythroid and megakaryocyte lineages.

In addition, we identified an Amerindian-population-specific variant at *MEF2C*, a known PLT locus. *MEF2C*, which encodes a MADS box transcription factor and is differentially expressed at various stages of hematopoiesis, is an important downstream target of stem cell leukemia for lineage-specific megakaryocyte development.[83] The Amerindian-ancestry-specific index SNP rs144261491 is distinct from the *MEF2C* European index SNP rs700585. rs144261491 is in LD with rs200572016 ($r^2 = 0.8$), which lies in a megakaryocyte-specific DNase hypersensitive site in the *MEF2C* antisense RNA and overlaps several transcription factor binding sites (GATA2, TAL1, and P300) in K562 cells. *MEF2C* and *MEF2C-AS1* are differentially expressed between erythroblasts and megakaryocytes.

In summary, we discovered and replicated three loci associated with PLT in Hispanic/Latino populations, as well as independent signals within two PLT-associated regions previously identified in populations of European descent. Several of these discovered PLT loci are prevalent among populations of Amerindian ancestry but rare or absent among populations of European or African ancestry. Amerindian-specific loci (e.g., *SLC16A11* [MIM: 615765]) for metabolic traits (diabetes and glycemic traits) similarly have been identified among other Hispanic/Latino populations.[84] Given the role of blood cells in pathogen invasion or defense, population-specific and/or rare variants might be expected to contribute to the regulation of genes relevant to quantitative blood cell phenotypes. The *ACTN1* and *MEF2C* alleles might have arisen by mutation among populations of Amerindian ancestry after the peopling of the Americas[85] and persisted as a result of local evolutionary selective pressure or genetic drift. Unlike the *SLC16A11* locus,[84] our PLT-associated loci did now show evidence of arising from introgression due to admixture with archaic humans. Taken together, our findings emphasize the importance and utility of performing genetic studies in populations with diverse ancestral backgrounds, including Hispanic/Latino populations.

## Supplemental Data

Supplemental Data include 6 figures and 12 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2015.12.003.

## Acknowledgments

We thank the participants and staff of the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). The baseline examination of HCHS/SOL was 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 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 (NIDDK), National Institute of Neurological Disorders and Stroke, and NIH Office of Dietary Supplements additionally contributed funding to HCHS/SOL. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Additional analysis support was provided by 1R01DK101855-01 and 13GRNT16490017. Genotyping was also supported by National Center for Advancing Translational Sciences UL1TR000124 and NIDDK DK063491 to the Southern California Diabetes Endocrinology Research Center. Additional support for rs117672662 functional studies was provided by R01 DK072193. This research was also supported in part by the Intramural Research Program of

the NIDDK, contract no. HHSB268201200054C, and Illumina. S.R.B. was supported by R01-GM110068. We thank Dr. Nick Patterson and the Simons Genome Diversity Project for kindly providing Australo-Melanesian sequence data. The Mount Sinai IPM Biobank Program is supported by the Andrea and Charles Bronfman Philanthropies.

Received: October 1, 2015
Accepted: December 7, 2015
Published: January 21, 2016

## Web Resources

The URLs for data presented herein are as follows:

HapMap recombination hotspots, ftp://ftp.hapmap.org/hapmap/recombination/2008-03_rel22_B36/rates/
OMIM, http://www.omim.org
UCSC Genome Browser, https://genome.ucsc.edu/

## References

1. Freson, K., Wijgaerts, A., and van Geet, C. (2014). Update on the causes of platelet disorders and functional consequences. Int. J. Lab. Hematol. *36*, 313–325.
2. 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.
3. Nürnberg, S.T., Rendon, A., Smethurst, P.A., Paul, D.S., Voss, K., Thon, J.N., Lloyd-Jones, H., Sambrook, J.G., Tijssen, M.R., Italiano, J.E., Jr., et al.; HaemGen Consortium (2012). A GWAS sequence variant for platelet volume marks an alternative DNM3 promoter in megakaryocytes near a MEIS1 binding site. Blood *120*, 4859–4868.
4. Msaouel, P., Lam, A.P., Gundabolu, K., Chrysofakis, G., Yu, Y., Mantzaris, I., Friedman, E., and Verma, A. (2014). Abnormal platelet count is an independent predictor of mortality in the elderly and is influenced by ethnicity. Haematologica *99*, 930–936.
5. Shameer, K., Denny, J.C., Ding, K., Jouni, H., Crosslin, D.R., de Andrade, M., Chute, C.G., Peissig, P., Pacheco, J.A., Li, R., et al. (2014). A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. Hum. Genet. *133*, 95–109.
6. Segal, J.B., and Moliterno, A.R. (2006). Platelet counts differ by sex, ethnicity, and age in the United States. Ann. Epidemiol. *16*, 123–130.
7. Bain, B.J. (1996). Ethnic and sex differences in the total and differential white cell count and platelet count. J. Clin. Pathol. *49*, 664–666.
8. Biino, G., Balduini, C.L., Casula, L., Cavallo, P., Vaccargiu, S., Parracciani, D., Serra, D., Portas, L., Murgia, F., and Pirastu, M. (2011). Analysis of 12,517 inhabitants of a Sardinian geographic isolate reveals that predispositions to thrombocytopenia and thrombocytosis are inherited traits. Haematologica *96*, 96–101.
9. Buckley, M.F., James, J.W., Brown, D.E., Whyte, G.S., Dean, M.G., Chesterman, C.N., and Donald, J.A. (2000). A novel approach to the assessment of variations in the human platelet count. Thromb. Haemost. *83*, 480–484.
10. 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.
11. Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y., and Kamatani, N. (2010). Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat. Genet. *42*, 210–215.
12. 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.
13. Qayyum, R., Snively, B.M., Ziv, E., Nalls, M.A., Liu, Y., Tang, W., Yanek, L.R., Lange, L., Evans, M.K., Ganesh, S., et al. (2012). A meta-analysis and genome-wide association study of platelet count and mean platelet volume in african americans. PLoS Genet. *8*, e1002491.
14. Soranzo, N., Spector, T.D., Mangino, M., Kühnel, B., Rendon, A., Teumer, A., Willenborg, C., Wright, B., Chen, L., Li, M., et al. (2009). A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat. Genet. *41*, 1182–1190.
15. Witkop, C.J., Nuñez Babcock, M., Rao, G.H., Gaudier, F., Summers, C.G., Shanahan, F., Harmon, K.R., Townsend, D., Sedano, H.O., King, R.A., et al. (1990). Albinism and Hermansky-Pudlak syndrome in Puerto Rico. Bol. Asoc. Med. P. R. *82*, 333–339.
16. Lavange, L.M., Kalsbeek, W.D., Sorlie, P.D., Avilés-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.
17. Sorlie, P.D., Avilés-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.
18. Rosenberg, N.A., Li, L.M., Ward, R., and Pritchard, J.K. (2003). Informativeness of genetic markers for inference of ancestry. Am. J. Hum. Genet. *73*, 1402–1422.
19. 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.; GENEVA Investigators (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. Genet. Epidemiol. *34*, 591–602.
20. Laurie, C.C., Laurie, C.A., Rice, K., Doheny, K.F., Zelnick, L.R., McHugh, C.P., Ling, H., Hetrick, K.N., Pugh, E.W., Amos, C., et al. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat. Genet. *44*, 642–650.
21. Conomos, M.P., Reiner, A.P., Weir, B.S., and Thornton, T.A. (2015). Model-free estimation of recent genetic relatedness. Am. J. Hum. Genet. *98*, 127–148.
22. Conomos, M.P., Miller, M.B., and Thornton, T.A. (2015). Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet. Epidemiol. *39*, 276–293.
23. 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.; 1000 Genomes Project Consortium (2012).

An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
24. Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. G3 (Bethesda) *1*, 457–470.
25. 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.
26. 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.
27. Conomos, M.P. (2014). Inferring, estimating and accounting for population and pedigree structure in genetic analyses. PhD thesis (University of Washington).
28. Conomos, M.P., Laurie, C.A., Stilp, A.M., Gogarten, S.M., McHugh, C.P., Nelson, S.C., Sofer, T., Fernández-Rhodes, L., Justice, A.E., Graff, M., et al. (2015). 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.
29. 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.
30. Clopper, C.J., and Pearson, E.S. (1934). The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika *26*, 404–413.
31. R Core Development Team (2014). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
32. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.
33. Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226.
34. Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L., Salzano, F.M., Patterson, N., and Reich, D. (2015). Genetic evidence for two founding populations of the Americas. Nature *525*, 104–108.
35. 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.
36. 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.
37. 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.
38. 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.
39. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
40. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288.
41. Cavalli-Sforza, L.L. (2005). The Human Genome Diversity Project: past, present and future. Nat. Rev. Genet. *6*, 333–340.
42. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., et al.; International HapMap 3 Consortium (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.
43. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097.
44. Heller, R., Bogomolov, M., and Benjamini, Y. (2014). Deciding whether follow-up studies have replicated findings in a preliminary large-scale omics study. Proc. Natl. Acad. Sci. USA *111*, 16262–16267.
45. 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.
46. Tijssen, M.R., Cvejic, A., Joshi, A., Hannah, R.L., Ferreira, R., Forrai, A., Bellissimo, D.C., Oram, S.H., Smethurst, P.A., Wilson, N.K., et al. (2011). Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. Dev. Cell *20*, 597–609.
47. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
48. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.
49. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. *40*, D930–D934.
50. Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. *32*, D91–D94.
51. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
52. Ritchie, G.R., Dunham, I., Zeggini, E., and Flicek, P. (2014). Functional annotation of noncoding sequence variants. Nat. Methods *11*, 294–296.
53. Lee, D., Gorkin, D.U., Baker, M., Strober, B.J., Asoni, A.L., McCallion, A.S., and Beer, M.A. (2015). A method to predict the impact of regulatory variants from DNA sequence. Nat. Genet. *47*, 955–961.
54. Knowler, W.C., Coresh, J., Elston, R.C., Freedman, B.I., Iyengar, S.K., Kimmel, P.L., Olson, J.M., Plaetke, R., Sedor, J.R.,

and Seldin, M.F.; Family Investigation of Nephropathy and Diabetes Research Group (2005). The Family Investigation of Nephropathy and Diabetes (FIND): design and methods. J. Diabetes Complications *19*, 1–9.
55. Tian, C., Hinds, D.A., Shigeta, R., Adler, S.G., Lee, A., Pahl, M.V., Silva, G., Belmont, J.W., Hanson, R.L., Knowler, W.C., et al. (2007). A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. Am. J. Hum. Genet. *80*, 1014–1023.
56. Fogarty, M.P., Cannon, M.E., Vadlamudi, S., Gaulton, K.J., and Mohlke, K.L. (2014). Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. PLoS Genet. *10*, e1004633.
57. Zuk, O., Hechter, E., Sunyaev, S.R., and Lander, E.S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. USA *109*, 1193–1198.
58. Heller, R., Bogomolov, M., Benjamini, Y., and Sofer, T. (2015). Testing for replicability in a follow-up study when the primary study hypotheses are two-sided. arXiv, arXiv:1503.02278.
59. Martens, J.H., and Stunnenberg, H.G. (2013). BLUEPRINT: mapping human blood cell epigenomes. Haematologica *98*, 1487–1489.
60. Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S., et al.; FANTOM consortium (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. *16*, 22.
61. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. *45*, 1238–1243.
62. Guéguen, P., Rouault, K., Chen, J.M., Raguénès, O., Fichou, Y., Hardy, E., Gobin, E., Pan-Petesch, B., Kerbiriou, M., Trouvé, P., et al. (2013). A missense mutation in the alpha-actinin 1 gene (ACTN1) is the cause of autosomal dominant macrothrombocytopenia in a large French family. PLoS ONE *8*, e74728.
63. Kunishima, S., Okuno, Y., Yoshida, K., Shiraishi, Y., Sanada, M., Muramatsu, H., Chiba, K., Tanaka, H., Miyazaki, K., Sakai, M., et al. (2013). ACTN1 mutations cause congenital macrothrombocytopenia. Am. J. Hum. Genet. *92*, 431–438.
64. Bottega, R., Marconi, C., Faleschini, M., Baj, G., Cagioni, C., Pecci, A., Pippucci, T., Ramenghi, U., Pardini, S., Ngu, L., et al. (2015). ACTN1-related thrombocytopenia: identification of novel families for phenotypic characterization. Blood *125*, 869–872.
65. Westbury, S.K., Turro, E., Greene, D., Lentaigne, C., Kelly, A.M., Bariana, T.K., Simeoni, I., Pillois, X., Attwood, A., Austin, S., et al.; BRIDGE-BPD Consortium (2015). Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. Genome Med. *7*, 36.
66. Hamill, K.J., Hiroyasu, S., Colburn, Z.T., Ventrella, R.V., Hopkinson, S.B., Skalli, O., and Jones, J.C. (2015). Alpha actinin-1 regulates cell-matrix adhesion organization in keratinocytes: consequences for skin cell motility. J. Invest. Dermatol. *135*, 1043–1052.
67. Raslova, H., Kauffmann, A., Sekkaï, D., Ripoche, H., Larbret, F., Robert, T., Tronik Le Roux, D., Kroemer, G., Debili, N., Dessen, P., et al. (2007). Interrelation between polyploidization and megakaryocyte differentiation: a gene profiling approach. Blood *109*, 3225–3234.
68. Thon, J.N., Macleod, H., Begonja, A.J., Zhu, J., Lee, K.C., Mogilner, A., Hartwig, J.H., and Italiano, J.E., Jr. (2012). Microtubule and cortical forces determine platelet size during vascular platelet production. Nat. Commun. *3*, 852.
69. Alharbi, R.A., Pettengell, R., Pandha, H.S., and Morgan, R. (2013). The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia *27*, 1000–1008.
70. Magnusson, M., Brun, A.C., Miyake, N., Larsson, J., Ehinger, M., Bjornsson, J.M., Wutz, A., Sigvardsson, M., and Karlsson, S. (2007). HOXA10 is a critical regulator for hematopoietic stem cells and erythroid/megakaryocyte development. Blood *109*, 3687–3696.
71. Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257–260.
72. Shivdasani, R.A., Fujiwara, Y., McDevitt, M.A., and Orkin, S.H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. EMBO J. *16*, 3965–3973.
73. Wang, C.C., Su, K.Y., Chen, H.Y., Chang, S.Y., Shen, C.F., Hsieh, C.H., Hong, Q.S., Chiang, C.C., Chang, G.C., Yu, S.L., and Chen, J.J. (2015). HOXA5 inhibits metastasis via regulating cytoskeletal remodelling and associates with prolonged survival in non-small-cell lung carcinoma. PLoS ONE *10*, e0124191.
74. Ablain, J., Durand, E.M., Yang, S., Zhou, Y., and Zon, L.I. (2015). A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. Dev. Cell *32*, 756–764.
75. Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L., et al. (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science *342*, 253–257.
76. Potter, M.D., Buijs, A., Kreider, B., van Rompaey, L., and Grosveld, G.C. (2000). Identification and characterization of a new human ETS-family transcription factor, TEL2, that is expressed in hematopoietic tissues and can associate with TEL1/ETV6. Blood *95*, 3341–3348.
77. Gu, X., Shin, B.H., Akbarali, Y., Weiss, A., Boltax, J., Oettgen, P., and Libermann, T.A. (2001). Tel-2 is a novel transcriptional repressor related to the Ets factor Tel/ETV-6. J. Biol. Chem. *276*, 9421–9436.
78. Wang, Q., Dong, S., Yao, H., Wen, L., Qiu, H., Qin, L., Ma, L., and Chen, S. (2014). ETV6 mutation in a cohort of 970 patients with hematologic malignancies. Haematologica *99*, e176–e178.
79. Zhang, M.Y., Churpek, J.E., Keel, S.B., Walsh, T., Lee, M.K., Loeb, K.R., Gulsuner, S., Pritchard, C.C., Sanchez-Bonilla, M., Delrow, J.J., et al. (2015). Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy. Nat. Genet. *47*, 180–185.
80. Noetzli, L., Lo, R.W., Lee-Sherick, A.B., Callaghan, M., Noris, P., Savoia, A., Rajpurkar, M., Jones, K., Gowan, K., Balduini, C.L., et al. (2015). Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. Nat. Genet. *47*, 535–538.
81. Topka, S., Vijai, J., Walsh, M.F., Jacobs, L., Maria, A., Villano, D., Gaddam, P., Wu, G., McGee, R.B., Quinn, E., et al. (2015). Germline ETV6 Mutations Confer Susceptibility to Acute Lymphoblastic Leukemia and Thrombocytopenia. PLoS Genet. *11*, e1005262.

82. Woo, A.J., Wieland, K., Huang, H., Akie, T.E., Piers, T., Kim, J., and Cantor, A.B. (2013). Developmental differences in IFN signaling affect GATA1s-induced megakaryocyte hyperproliferation. J. Clin. Invest. Published online July 1, 2013. http://dx.doi.org/10.1172/JCI40609.
83. Gekas, C., Rhodes, K.E., Gereige, L.M., Helgadottir, H., Ferrari, R., Kurdistani, S.K., Montecino-Rodriguez, E., Bassel-Duby, R., Olson, E., Krivtsov, A.V., et al. (2009). Mef2C is a lineage-restricted target of Scl/Tal1 and regulates megakaryopoiesis and B-cell homeostasis. Blood *113*, 3461–3471.
84. Williams, A.L., Jacobs, S.B., Moreno-Macías, H., Huerta-Chagoya, A., Churchhouse, C., Márquez-Luna, C., García-Ortíz, H., Gómez-Vázquez, M.J., Burtt, N.P., Aguilar-Salinas, C.A., et al.; SIGMA Type 2 Diabetes Consortium (2014). Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature *506*, 97–101.
85. Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature *488*, 370–374.

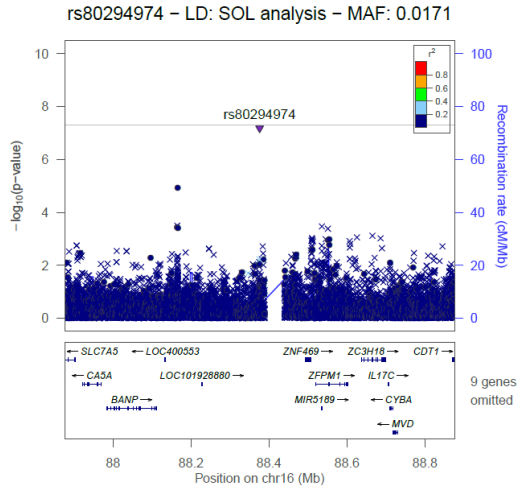
**The American Journal of Human Genetics**

**Supplemental Data**

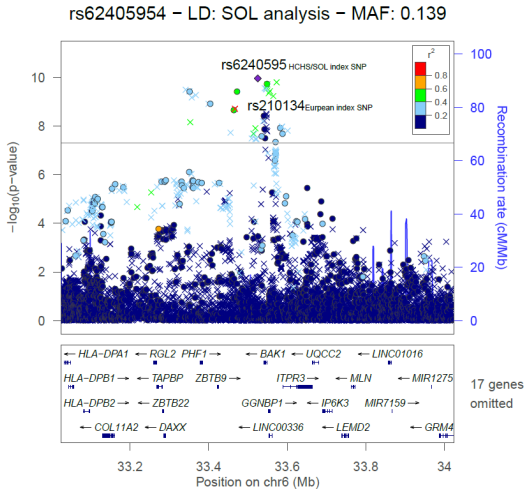
# Genome-wide Association Study of Platelet Count Identifies Ancestry-Specific Loci in Hispanic/Latino Americans

**Ursula M. Schick, Deepti Jain, Chani J. Hodonsky, Jean V. Morrison, James P. Davis, Lisa Brown, Tamar Sofer, Matthew P. Conomos, Claudia Schurmann, Caitlin P. McHugh, Sarah C. Nelson, Swarooparani Vadlamudi, Adrienne Stilp, Anna Plantinga, Leslie Baier, Stephanie A. Bien, Stephanie M. Gogarten, Cecelia A. Laurie, Kent D. Taylor, Yongmei Liu, Paul L. Auer Nora Franceschini, Adam Szpiro, Ken Rice, Kathleen F. Kerr, Jerome I. Rotter, Robert L. Hanson, George Papanicolaou, Stephen S. Rich, Ruth J.F. Loos, Brian L. Browning, Sharon R. Browning, Bruce S. Weir, Cathy C. Laurie, Karen L. Mohlke, Kari E. North, Timothy A. Thornton, and Alex P. Reiner**

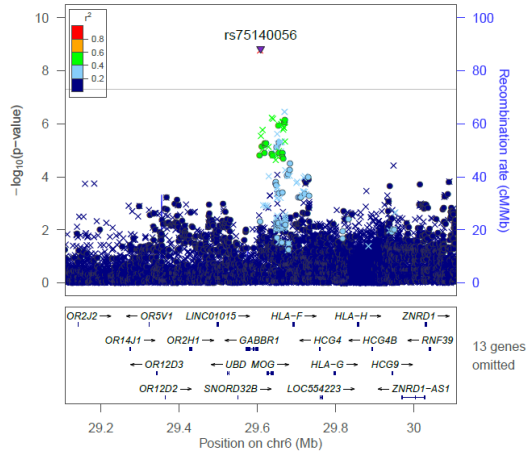
**A.**


rs80294974 − LD: SOL analysis − MAF: 0.0171
rs80294974
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
SLC7A5 LOC400553 ZNF469 ZC3H18 CDT1
CA5A LOC101928880 ZFPM1 IL17C
BANP MIR5189 CYBA
MVD
9 genes omitted
88 88.2 88.4 88.6 88.8
Position on chr16 (Mb)

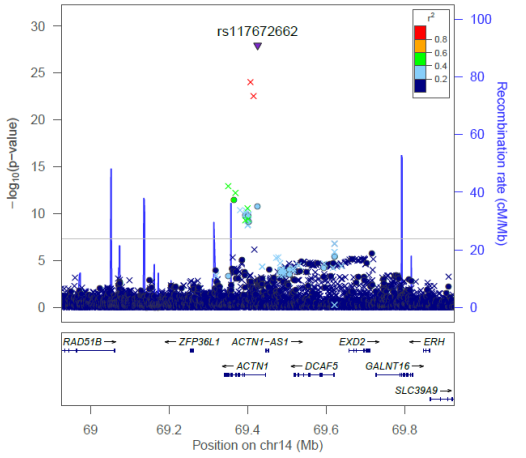
**B.**


rs62405954 − LD: SOL analysis − MAF: 0.139
rs6240595 HCHS/SOL index SNP
rs210134 European index SNP
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
HLA−DPA1 RGL2 PHF1 BAK1 UQCC2 LINC01016
HLA−DPB1 TAPBP ZBTB9 ITPR3 MLN MIR1275
HLA−DPB2 ZBTB22 GGNBP1 IP6K3 MIR7159
COL11A2 DAXX LINC00336 LEMD2 GRM4
17 genes omitted
33.2 33.4 33.6 33.8 34
Position on chr6 (Mb)

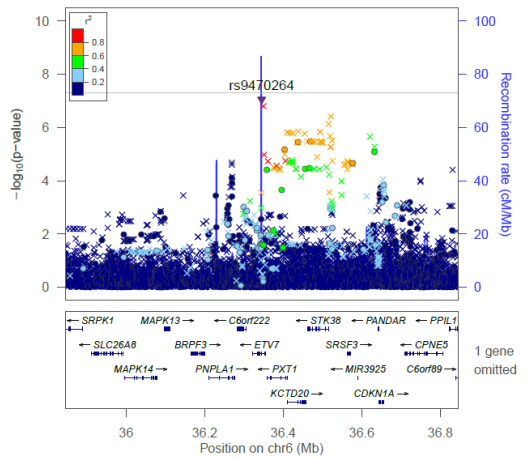
**C.**


chr6:29608184 − LD: SOL analysis − MAF: 0.394
rs75140056
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
OR2J2 OR5V1 LINC01015 HLA−F HLA−H ZNRD1
OR14J1 OR2H1 GABBR1 HCG4 HCG4B RNF39
OR12D3 UBD MOG HLA−G HCG9
OR12D2 SNORD32B LOC554223 ZNRD1−AS1
13 genes omitted
29.2 29.4 29.6 29.8 30
Position on chr6 (Mb)

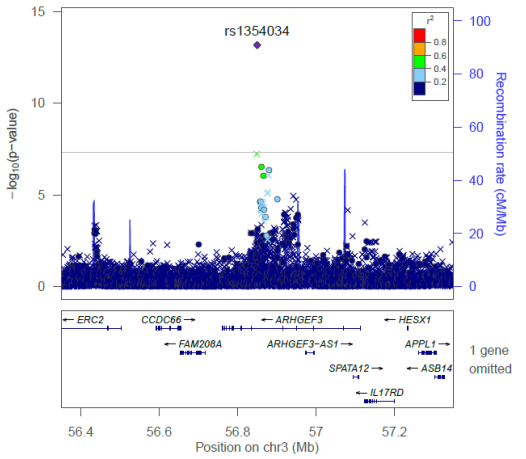
**D.**


rs117672662 − LD: SOL analysis − MAF: 0.0596
rs117672662
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
RAD51B ZFP36L1 ACTN1−AS1 EXD2 ERH
ACTN1 DCAF5 GALNT16
SLC39A9
69 69.2 69.4 69.6 69.8
Position on chr14 (Mb)

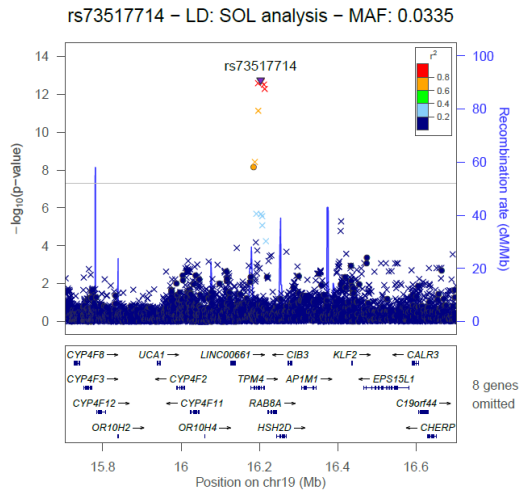
**E.**


rs9470264 − LD: SOL analysis − MAF: 0.201
rs9470264
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
SRPK1 MAPK13 C6orf222 STK38 PANDAR PPIL1
SLC26A8 BRPF3 ETV7 SRSF3 CPNE5
MAPK14 PNPLA1 PXT1 MIR3925 C6orf89
KCTD20 CDKN1A
1 gene omitted
36 36.2 36.4 36.6 36.8
Position on chr6 (Mb)

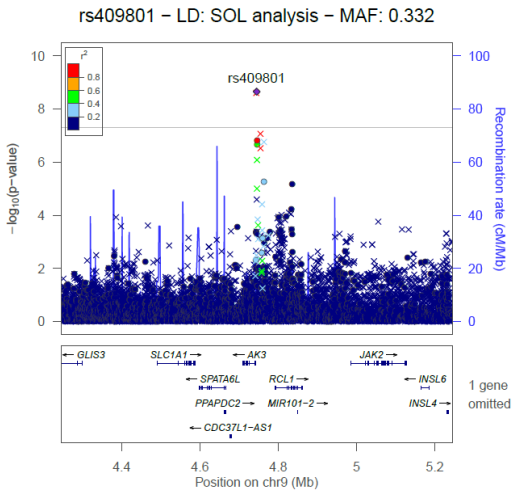
**F.**


rs1354034 − LD: SOL analysis − MAF: 0.427
rs1354034
−log10(p−value)
Recombination rate (cM/Mb)
r²
0.8
0.6
0.4
0.2
ERC2 CCDC66 ARHGEF3 HESX1
FAM208A ARHGEF3−AS1 APPL1
SPATA12 ASB14
IL17RD
1 gene omitted
56.4 56.6 56.8 57 57.2
Position on chr3 (Mb)

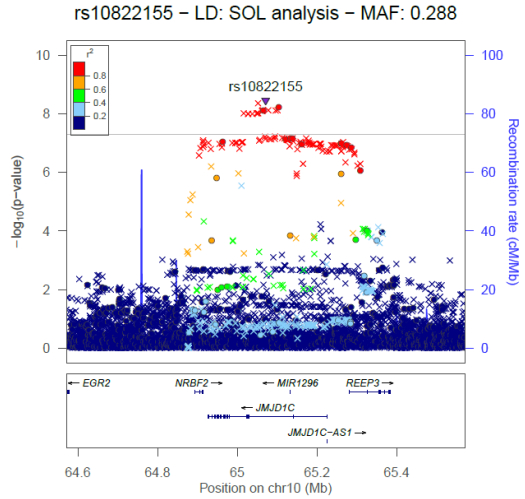
**G.**


rs73517714 – LD: SOL analysis – MAF: 0.0335
rs73517714
CYP4F8 UCA1 LINC00661 CIB3 KLF2 CALR3
CYP4F3 CYP4F2 TPM4 AP1M1 EPS15L1
CYP4F12 CYP4F11 RAB8A C19orf44
OR10H2 OR10H4 HSH2D CHERP
8 genes omitted
Position on chr19 (Mb)
Recombination rate (cM/Mb)
−log10(p-value)

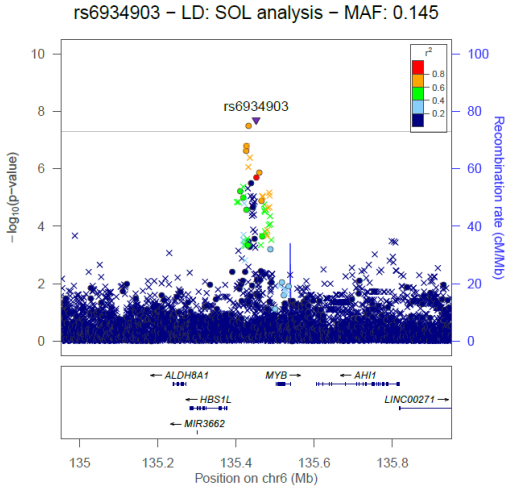
**H.**


rs409801 – LD: SOL analysis – MAF: 0.332
rs409801
GLIS3 SLC1A1 AK3 JAK2
SPATA6L RCL1 INSL6
PPAPDC2 MIR101-2 INSL4
CDC37L1-AS1
1 gene omitted
Position on chr9 (Mb)
Recombination rate (cM/Mb)
−log10(p-value)

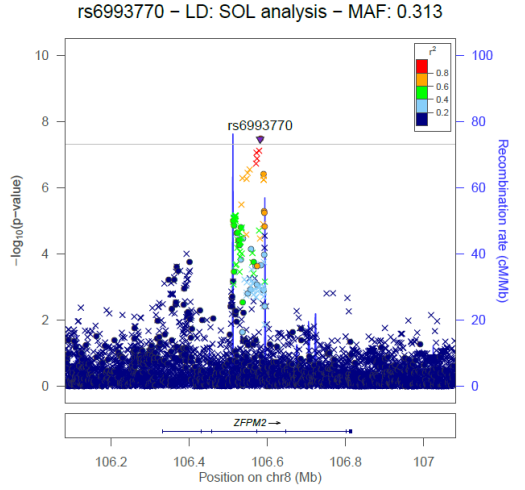
**I. .**


rs10822155 – LD: SOL analysis – MAF: 0.288
rs10822155
EGR2 NRBF2 MIR1296 REEP3
JMJD1C
JMJD1C-AS1
Position on chr10 (Mb)
Recombination rate (cM/Mb)
−log10(p-value)

**J.**


rs6934903 – LD: SOL analysis – MAF: 0.145
rs6934903
ALDH8A1 MYB AHI1
HBS1L LINC00271
MIR3662
Position on chr6 (Mb)
Recombination rate (cM/Mb)
−log10(p-value)

**K.**


rs6993770 – LD: SOL analysis – MAF: 0.313
rs6993770
ZFPM2
Position on chr8 (Mb)
Recombination rate (cM/Mb)
−log10(p-value)

**L.**

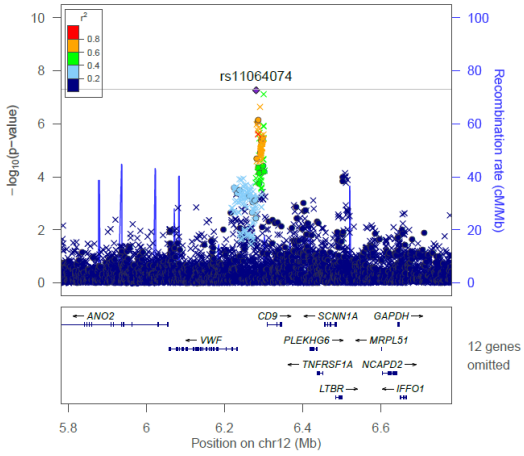
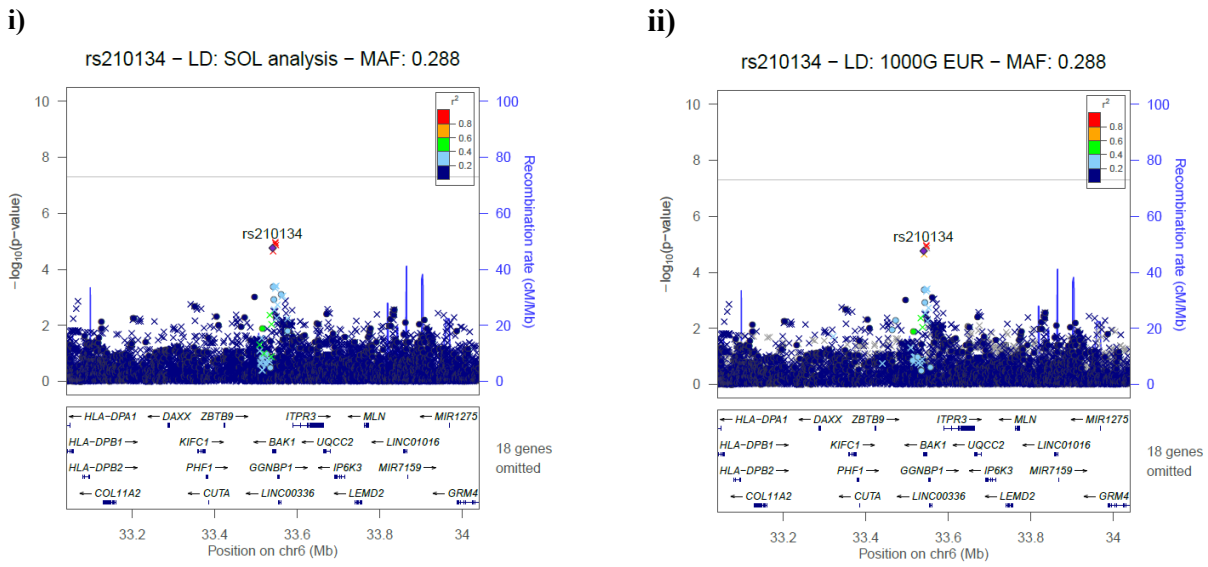

rs11064074 – LD: SOL analysis – MAF: 0.289
rs11064074
ANO2 CD9 SCNN1A GAPDH
VWF PLEKHG6 MRPL51
TNFRSF1A NCAPD2
LTBR IFFO1
12 genes omitted
Position on chr12 (Mb)
Recombination rate (cM/Mb)
−log10(p-value)

**Figure S1**: LocusZoom plots of loci that have discovery p-values less than $1x10^{-7}$. In each, the top panel reflects the main PLT GWAS analysis results. The LD estimates were calculated in the SOL discovery sample and are presented with respect to the lowest P-value SNP (reference). The genotyped reference SNP is denoted by a filled diamond, imputed reference SNP is denoted by a filled triangle, other imputed SNPs are denoted by a cross, and other genotyped SNPs are denoted by a filled circle. Recombination hotspots are indicated by the blue lines and scaled according to the y-axis on the right side of the plot. The dashed line indicates the nominal significance threshold p-value $\leq 5 \times 10^{-8}$. The bottom panel shows the genes and their orientation for each region. A. *BANP-ZFPM1* with rs80294974 as index variant, B. *ZBTB9-BAK1* with rs62405954 as index variant, C. *GABBR1-MOG* with rs75140056 as index variant, D. *ACTN1* with rs117672662 as index variant, E. *ETV7* with rs9470264 as index variant, F. *ARGHEF3* with rs1354034 as index variant, G. *TPM4* with rs73517714 as index variant, H. *AK3-RCL1* with rs409801 as index variant, I. *JMJD1C* with rs10822155 as index variant, J. *HBS1L-MYB* with rs6934903 as index variants, K. *ZFPM2* with rs6993770 as index variant, L. *VWF-CD9* with rs11064074 as index variant.

**A.**




**B.**

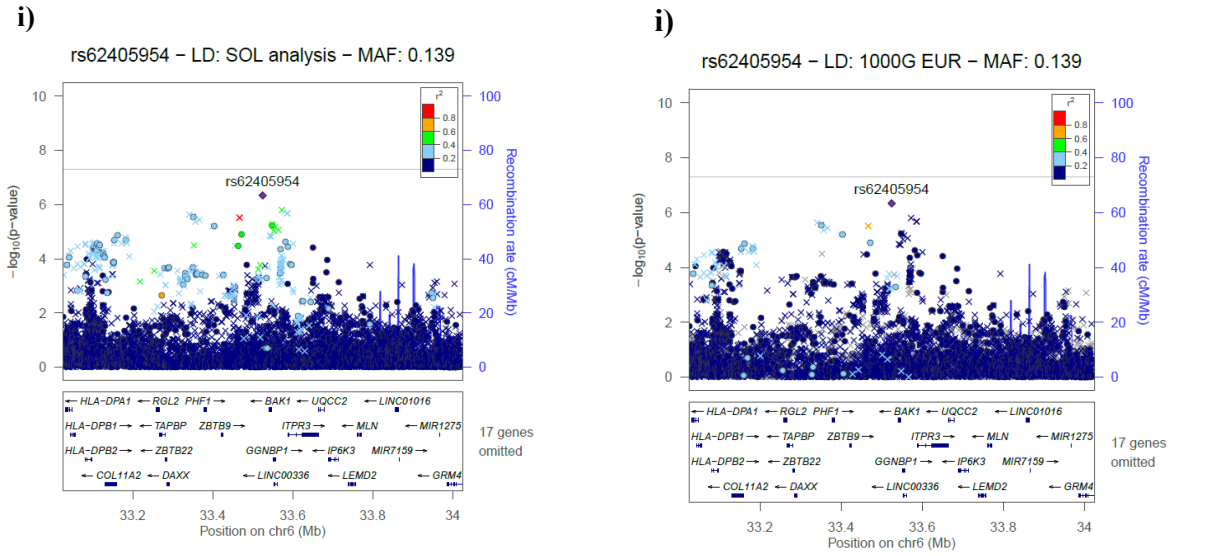
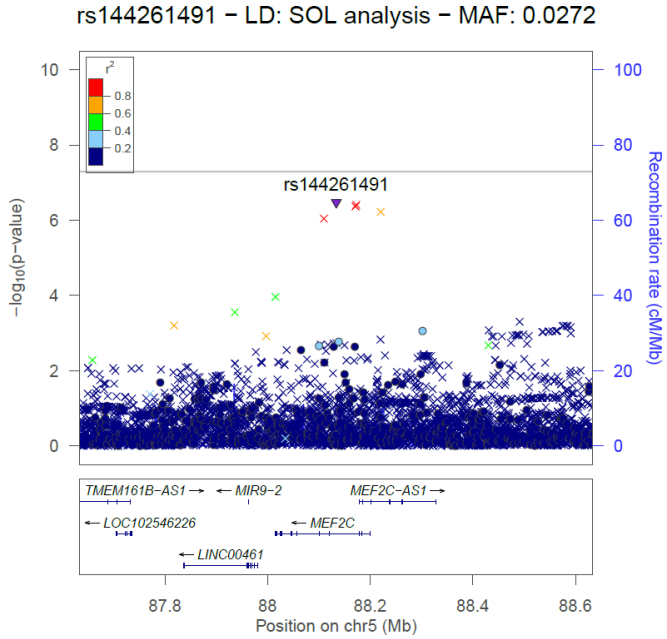
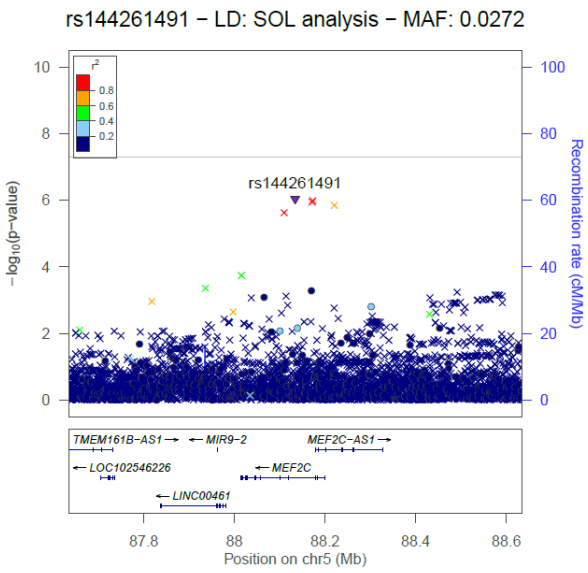



**Figure S2. LocusZoom plots of conditional analysis at *ZBTB9-BAK1* locus.** In each, the top panel reflects the conditional analysis results. The LD estimates with respect to the indicated reference SNP at each locus are color coded based on the scale indicated in each plot. As indicated, the LD estimates are either derived using HCHS/SOL population or the 1000 genomes EUR super-population. The genotyped reference SNP is denoted by a filled diamond, imputed reference SNP is denoted by a filled triangle, other imputed SNPs are denoted by a cross, and other genotyped SNPs are denoted by a filled circle. Recombination hotspots are indicated by the blue lines. The dashed line indicates the nominal significance threshold p-value $\leq 5 \times 10^{-8}$.The bottom panel shows the genes and their orientation for each region. **A. C**onditioned on HCHS/SOL index SNP rs62405954 and color coded by i) LD estimates from I HCHS/SOL population, and ii) LD estimates from 1000genomes EUR super-population. B) **B.** Conditioned on the European GWAS index SNP rs210134 and color coded by i) LD estimates from HCHS/SOL population, and ii) LD estimates from 1000genomes EUR super-population.

**A.**




**B.**




**C.**

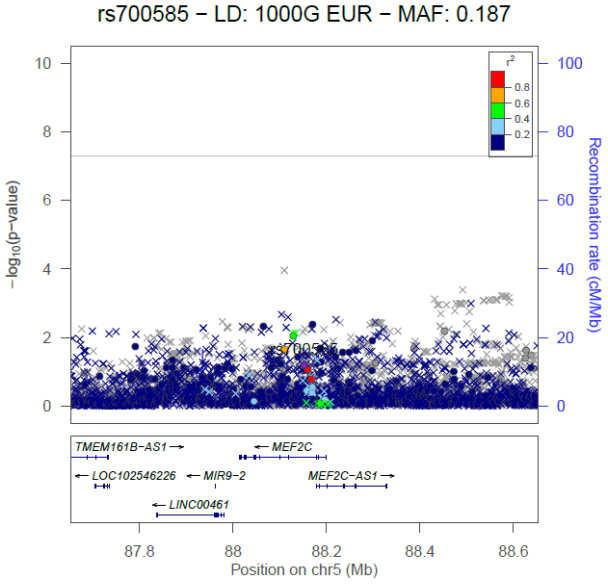
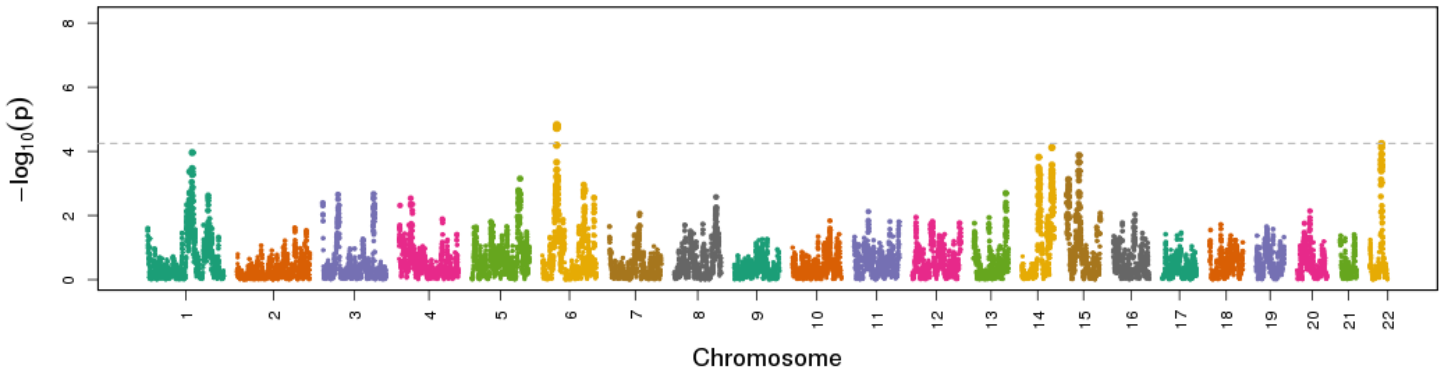



**Figure S3. LocusZoom plots of main and conditional analysis at *MEF2C* locus.** In each, the top panel reflects the main PLT GWAS or conditional analysis results, as indicated. The LD estimates with respect to the indicated reference SNP at each locus are color coded based on the scale indicated in each plot. As noted, the LD estimates are either derived using HCHS/SOL population or the 1000 genomes EUR super-population. The genotyped reference SNP is denoted by a filled diamond, imputed reference SNP is denoted by a filled triangle, other imputed SNPs are denoted by a cross and other genotyped SNPs are denoted by a filled circle. Recombination hotspots are indicated by the blue lines. The dashed line indicates the nominal significance threshold p-value $\leq 5 \times 10^{-8}$. The bottom panel shows the genes and their orientation for each region.**A.** MEF2C locus from the main PLT GWAS analysis color coded by LD estimates from HCHCS/SOL with reference to HCHS/SOL index SNP rs144261491. **B.** *MEF2C* locus conditioned on European GWAS index SNP rs700585 and color coded by LD estimates from HCHS/SOL population. **C**. *MEF2C* locus conditioned on HCHS/SOL index SNP rs144261491 and color coded LD estimates from 1000 genomes EUR super-population

**A.**




**B.**

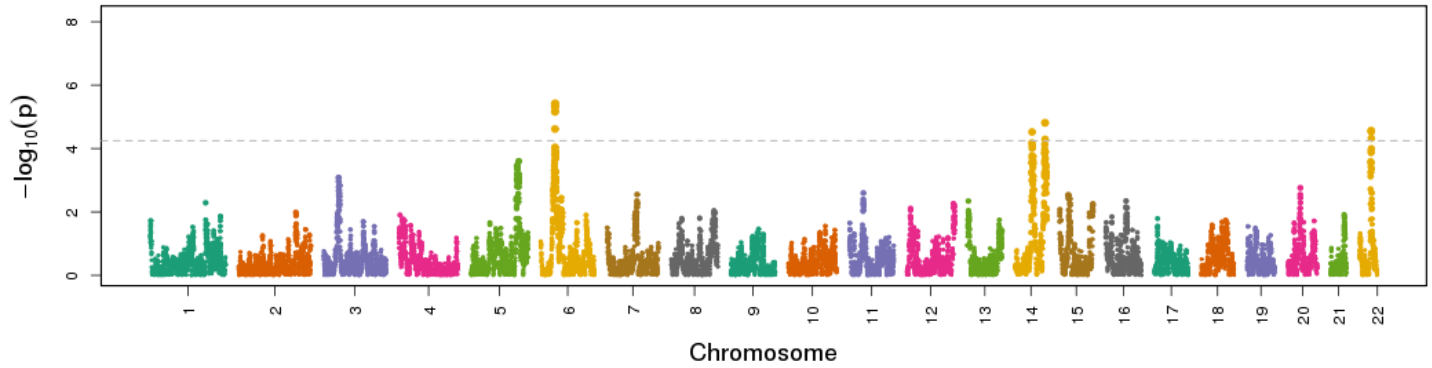



**Figure S4: A.** Genome-wide admixture scan with a joint test of all three local ancestry estimates (Amerindian, European, African) from RFMix with dashed line indicating the significance threshold p-value≤ $-\log_{10}(5.7 \times 10^{-5})$, B. Genome-wide admixture scan of Amerindian against any other ancestry from from RFMix with dashed line indicating the significance threshold p-value≤ $-\log_{10}(5.7 \times 10^{-5})$.

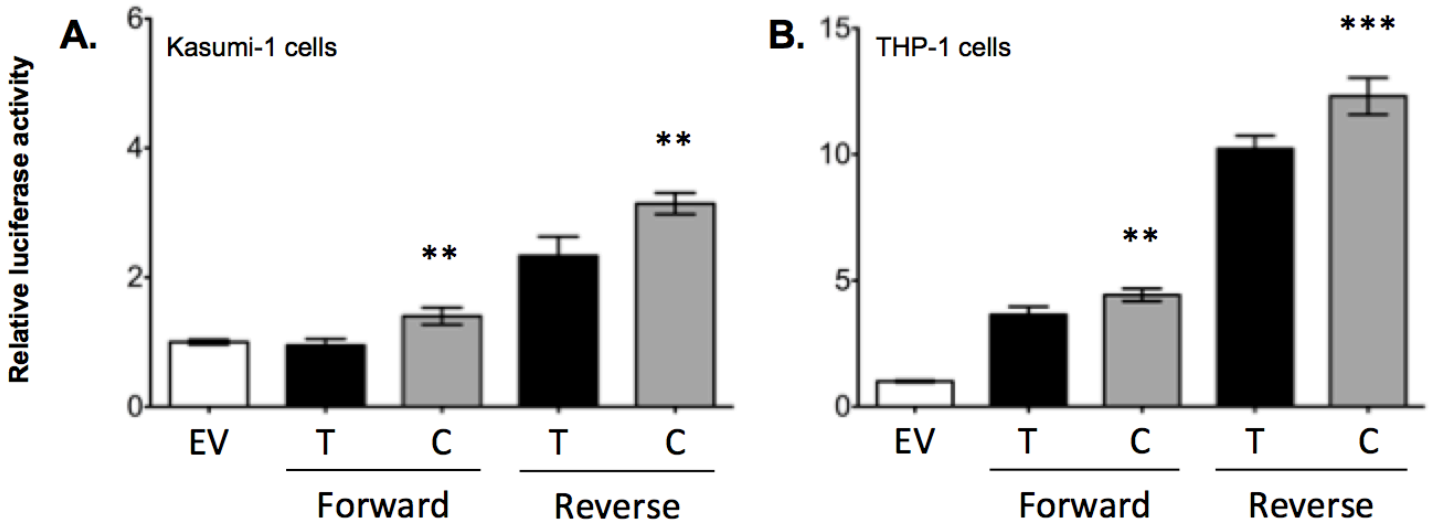



**Figure S5: Allelic differences in transcriptional activity of variant rs117672662**. A 186 bp PCR product approximately centered on rs117672662 was cloned in both orientations (Forward and Reverse) relative to a minimal promoter in a transcriptional reporter vector (pGL4.23). Clones were transfected into (A) Kasumi-1 cells: a myeloblast, cell line established from the peripheral blood of an acute myeloid leukemia patient and (B) THP-1 cells: a monocyte cell line, isolated from peripheral blood of an acute monocyte leukemia patient. The cells were incubated for 48 hrs, and luciferase and Renilla activity were measured in the cell lysate. Results are shown as relative luciferase/Renilla activity normalized to empty vector. Error bars represent the standard deviation of results for four clones. EV: Empty Vector

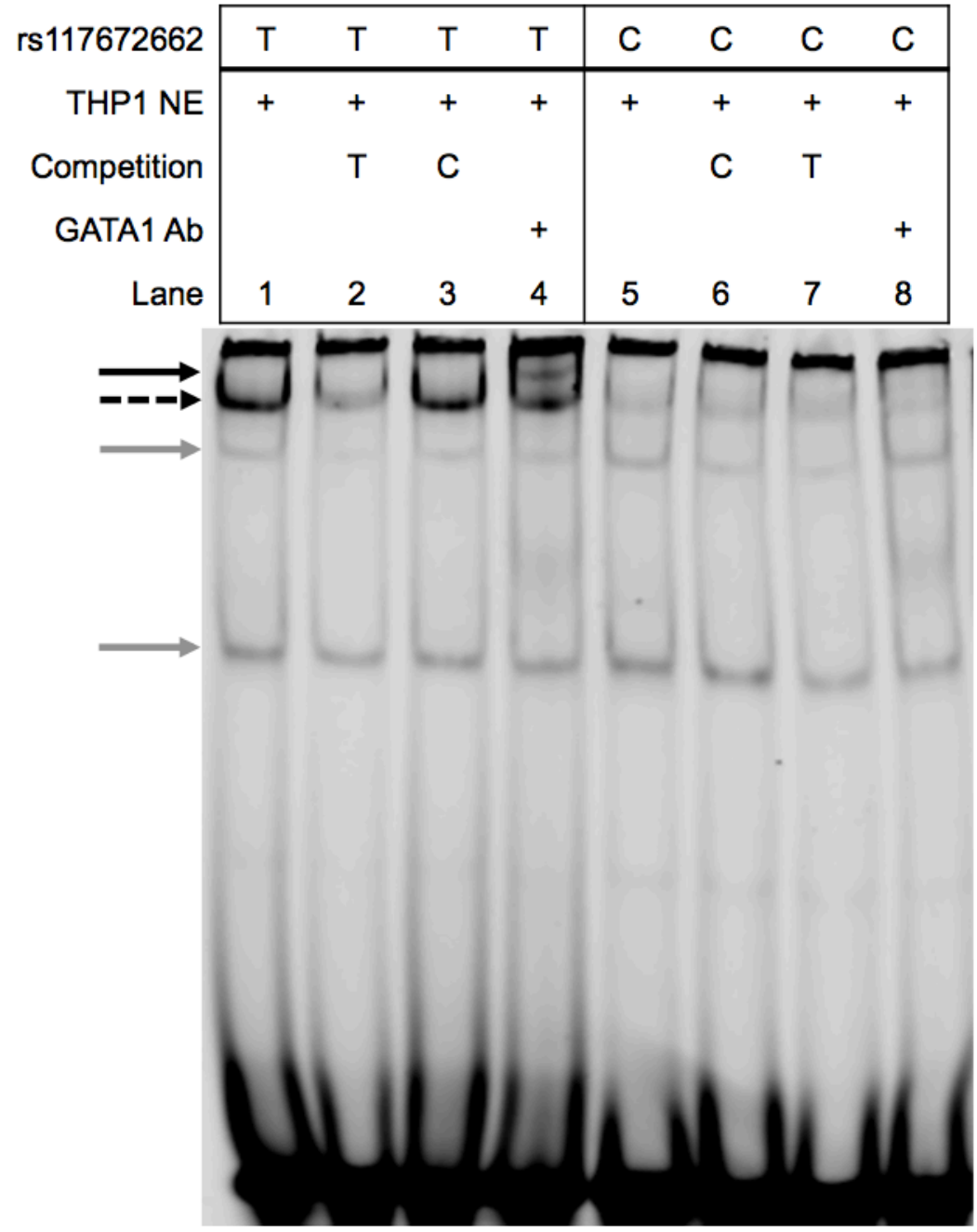



**Figure S6: Further support of allelic differences in protein binding at rs117672662**. EMSA using oligonucleotide probes containing different alleles at rs117672662 (T allele: lanes 1 to 4, C allele: lanes 5 to 8). Nuclear extracts from human monocyte THP-1 cells were incubated with IRDye-labeled double-stranded oligonucleotide probe alone (lanes 1, 5) or with 70-fold excess of unlabeled probe (lanes 2, 3, 6 and 7) or GATA1 antibodies (lanes 4 and 8). The dotted black arrow indicates probe-protein complexes, the solid black arrow indicates probe-protein-antibody complexes and the grey arrows indicate non-specific probe-protein binding complexes. The unlabeled T-allele probe shows stronger competition than the unlabeled C-allele probe. The faint band remaining in lane 2 may represent incomplete competition or another non-specific probe-protein complex.

**Table S1.** Demographic Characteristics of Discovery and Replication Cohorts

| Study | Phase | N | Study Design | Mean Platelet Count[a] (Standard Deviation) | % Female | Age[b] Mean (Standard Deviation) |
|---|---|---|---|---|---|---|
| HCHS/SOL | Discovery | 12941 | Population- and family-based | 252 (65) | 59 | 46 (14) |
| Bio*Me* | Replication | 2854 | Population-based, unrelated | 233 (71) | 62 | 59 (15) |
| WHI | Replication | 3534 | Population-based, unrelated | 245 (56) | 100 | 60 (7) |
| MESA | Replication | 782 | Population-based, unrelated | 234 (66) | 53 | 69 (9) |

[a]Platelet count unit = $10^9$/L, [b]Age = years.

**Table S2**. Variants reaching the genome-wide (P-value<$5x10^{-8}$) or suggestive (P-value< $1x10^{-7}$) significance in HCHS/SOL

| Novel or Known | Nearest Gene(s) (genomic context) | Chromosome: position (GRCh37/hg19) | rsID | Coded / alternative allele on (+) strand | HCHS/SOL Coded Allele Frequency | N | Beta (Standard Error) | P-value |
|---|---|---|---|---|---|---|---|---|
| Novel | *ACTN1* (intronic) | chr14:69425467 | rs117672662 | T/C | 0.94 | 12491 | 0.604(0.054) | $1.16x10^{-28}$ |
| Novel | *ZBTB9-BAK1* (intergenic) | chr6:33524820 | rs62405954 | T/C | 0.86 | 12491 | -0.239(0.037) | $1.10x10^{-10}$ |
| Novel | *GABBR1-MOG* (intergenic) | chr6:29608184 | rs75140056 | C/CAT | 0.39 | 12491 | -0.151(0.025) | $1.47x10^{-09}$ |
| Novel | *BANP-ZFPM1* (intergenic) | chr16:88376014 | rs80294974 | G/A | 0.98 | 12491 | 0.555(0.103) | $6.60x10^{-08}$ |
| Novel | *ETV7* (intronic) | chr6:36344980 | rs9470264 | G/A | 0.8 | 12491 | -0.181(0.034) | $8.80x10^{-08}$ |
| Known | *ARGHEF3* (intronic) | chr3:56849749 | rs1354034 | T/C | 0.57 | 12491 | -0.189(0.025) | $6.89x10^{-14}$ |
| Known | *TPM4* (intronic) | chr19:16203304 | rs73517714 | C/A | 0.97 | 12491 | 0.508(0.069) | $1.80x10^{-13}$ |
| Known | *AK3-RCL1* (integenic) | chr9:4744743 | rs409801 | T/C | 0.67 | 12491 | -0.163(0.027) | $2.21x10^{-09}$ |
| Known | *JMJD1C* (intronic) | chr10:65071215 | rs10822155 | C/A | 0.71 | 12491 | -0.168(0.028) | $3.50x10^{-09}$ |
| Known | *HBS1L-MYB* (intergenic) | chr6:135451564 | rs6934903 | T/A | 0.85 | 12491 | -0.198(0.035) | $2.02x10^{-08}$ |
| Known | *ZFPM2* (intronic) | chr8:106581528 | rs6993770 | A/T | 0.69 | 12490 | 0.148(0.027) | $3.43x10^{-08}$ |
| Known | *VWF-CD9* (intergenic) | chr12:6281039 | rs11064074 | C/T | 0.71 | 12491 | 0.150(0.028) | $5.38x10^{-08}$ |

**Table S3.** Results from the generalization of European[1-4], African-American[5], and Japanese[6] GWAS PLT-associated variants to HCHS/SOL Hispanic/Latinos. The line separates the SNPs that generalized (top panel) from those that failed to generalize (bottom panel).

| rsID | Nearest gene(s) (genomic context ) | Chr:Position (build 37-hg19) | Coded / Alt. Allele | Discovery | | | | | HCHS/SOL | | | | | Meta-analysis | | Ref. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Ancestry | N | MAF[a] | Beta (SE) | p-value | MAF | CAF | Beta (SE) | 2-sided p-value | r-value | Beta (SE) | p-value | |
| rs3811444 | *TRIM58* (nonsynonymous) | 1:248039450 | C/T | European | 27955 | NA | 3.35 (0.57) | 5.6E-09 | 0.28 | 0.72 | 3.45 (0.89) | 1.06E-04 | 5.58E-04 | 3.38 (0.48) | 2.57E-12 | [1] |
| rs1260326 | *GCKR* (nonsynonymous) | 2:27730939 | T/C | European | 54396 | NA | 2.33 (0.38) | 9.12E-10 | 0.34 | 0.34 | 2.53 (0.86) | 3.12E-03 | 9.35E-03 | 2.37 (0.35) | 1.06E-11 | [1] |
| rs625132 | *EHD3* (intronic) | 2:31482299 | G/A | European | 45217 | NA | 4.24 (0.57) | 9.15E-14 | 0.11 | 0.89 | 3.93 (1.25) | 1.67E-03 | 5.81E-03 | 4.18 (0.52) | 5.97E-16 | [1] |
| rs1354034 | *ARHGEF3* (intronic) | 3:56849749 | T/C | European | 6234 | 0.41 | -9.44 (1.61) | 4.35E-09 | 0.43 | 0.57 | -5.87 (0.8) | 2.44E-13 | 4.75E-04 | -6.58 (0.72) | 4.41E-20 | [2] |
| rs1354034 | *ARHGEF3* (intronic) | 3:56849749 | C/T | European | 13582 | 0.39 | 7.97 (0.79) | 6.00E-24 | 0.43 | 0.43 | 5.87 (0.8) | 2.44E-13 | 7.32E-13 | 6.94 (0.56) | 6.39E-35 | [3] |
| rs6141 | *THPO* (3' downstream) | 3:184090266 | T/C | Japanese | 14806 | 0.45 | 0.08 (0.01) | 5.38E-11 | 0.33 | 0.67 | 0.06 (0.01) | 3.61E-05 | 3.61E-05 | 0.07 (0.01) | 6.75E-14 | [6] |
| rs700585 | *MEF2C* (intronic) | 5:88152116 | C/T | European | 55469 | NA | 2.7 (0.44) | 9.86E-10 | 0.19 | 0.19 | 2.41 (1.01) | 1.70E-02 | 3.76E-02 | 2.66 (0.4) | 5.42E-11 | [1] |
| rs12526480 | *LRRC16A* (intronic) | 6:25533534 | G/T | African | 16388 | 0.31 | -4.39 (0.76) | 9.15E-29 | 0.36 | 0.36 | -2.6 (0.82) | 1.53E-03 | 2.54E-03 | -3.56 (0.56) | 1.66E-10 | [5] |
| rs210134 | *BAK1* (3' downstream) | 6:33540209 | A/G | African | 16388 | 0.29 | -6.16 (0.78) | 2.32E-15 | 0.29 | 0.29 | -4.94 (0.86) | 9.31E-09 | 9.31E-08 | -5.61 (0.58) | 2.79E-22 | [5] |
| rs5745568 | *BAK1* (5' upstream) | 6:33548394 | T/G | Japanese | 14806 | 0.23 | 0.09 (0.01) | 6.66E-11 | 0.24 | 0.24 | 0.10 (0.01) | 1.89E-10 | 7.50E-06 | 0.09 (0.01) | 1.48E-19 | [6] |
| rs4895441 | *HBS1L-MYB* (intergenic) | 6:135426573 | A/G | European | 13582 | 0.27 | -5.42 (0.88) | 9.00E-10 | 0.18 | 0.82 | -5.39 (1.02) | 1.24E-07 | 4.29E-05 | -5.41 (0.67) | 5.84E-16 | [3] |
| rs9494145 | *HBS1L-MYB* (intergenic) | 6:135432552 | C/T | African | 16388 | 0.07 | 8.19 (1.38) | 2.79E-09 | 0.16 | 0.16 | 5.88 (1.07) | 3.92E-08 | 9.81E-08 | 6.75 (0.85) | 1.48E-15 | [5] |
| rs342275 | *AF086203* (intronic) | 7:106359215 | C/T | European | 58571 | NA | 3.74 (0.36) | 5.57E-25 | 0.33 | 0.67 | 3.6 (0.86) | 2.82E-05 | 2.37E-04 | 3.72 (0.33) | 9.42E-29 | [1] |
| rs342293 | *AF086203* (intronic) | 7:106372219 | G/C | African | 16388 | 0.39 | -4.05 (0.72) | 1.58E-08 | 0.39 | 0.39 | -3.41 (0.82) | 2.92E-05 | 5.84E-05 | -3.77 (0.54) | 2.88E-12 | [5] |
| rs6993770 | *ZFPM2* (intronic) | 8:106581527 | A/T | European | 54960 | NA | 3.67 (0.44) | 4.30E-17 | 0.31 | 0.69 | 4.74 (0.85) | 2.87E-08 | 1.21E-06 | 3.89 (0.39) | 1.52E-23 | [1] |
| rs6995402 | *PLEC* (intronic) | 8:145005560 | C/T | European | 57593 | NA | 2.30 (0.37) | 5.09E-10 | 0.05 | 0.05 | 5.27 (2.24) | 1.88E-02 | 3.96E-02 | 2.38 (0.37) | 7.50E-11 | [1] |
| rs385893 | *AK3-RCL1* (intergenic) | 9:4763176 | T/C | Japanese | 14806 | 0.25 | -0.10 (0.01) | 2.95E-13 | 0.40 | 0.60 | -0.06 (0.01) | 5.44E-06 | 1.09E-05 | -0.08 (0.01) | 2.87E-16 | [6] |
| rs385893 | *AK3-RCL1* (intergenic) | 9:4763176 | C/T | European | 9316 | 0.44 | 6.26 (0.75) | 8.50E-17 | 0.40 | 0.40 | 3.74 (0.83) | 5.88E-06 | 1.18E-05 | 5.12 (0.56) | 3.62E-20 | [4] |
| rs423955 | *AK3-RCL1* (intergenic) | 9:4792339 | A/G | European | 13582 | 0.34 | 4.94 (0.81) | 1.0E-9 | 0.40 | 0.60 | 1.85 (0.82) | 2.33E-02 | 2.33E-02 | 3.41 (0.57) | 2.89E-09 | [3] |
| rs10761731 | *RCL1* (5' upstream) | 10:65027609 | T/A | European | 54344 | NA | 3.85 (0.38) | 2.02E-24 | 0.31 | 0.31 | 4.81 (0.89) | 6.44E-08 | 1.35E-06 | 4.00 (0.35) | 1.56E-30 | [1] |
| rs7896518 | *JMJD1C* (intronic) | 10:65104500 | G/A | African | 16388 | 0.32 | 5.18 (0.74) | 2.26E-212 | 0.30 | 0.30 | 4.89 (0.89) | 3.91E-08 | 9.81E-08 | 5.06 (0.57) | 5.82E-19 | [5] |
| rs505404 | *PSMD13* (intronic) | 11:243267 | G/T | European | 54642 | NA | 4.66 (0.45) | 7.44E-25 | 0.26 | 0.26 | 2.53 (0.91) | 5.68E-03 | 1.59E-02 | 4.24 (0.41) | 1.48E-25 | [1] |
| rs4938642 | *CBL* (intronic) | 11:119099905 | C/G | European | 56605 | NA | 4.73 (0.73) | 7.66E-11 | 0.05 | 0.05 | 5.57 (1.75) | 1.47E-03 | 5.62E-03 | 4.85 (0.67) | 4.89E-13 | [1] |
| rs7342306 | *VWF-CD9* (intergenic) | 12:6291092 | G/A | European | 55636 | NA | 2.53 (0.38) | 4.29E-11 | 0.25 | 0.75 | 4.81 (0.92) | 1.69E-07 | 7.75E-06 | 2.87 (0.35) | 5.51E-16 | [1] |
| rs4326844 | *COPZ1* (intronic) | 12:54736470 | A/G | European | 6234 | 0.45 | -8.68 (1.59) | 4.57E-08 | 0.27 | 0.27 | -2.81 (0.93) | 2.61E-03 | 2.61E-03 | -4.31 (0.8) | 7.95E-08 | [2] |
| rs941207 | *BAZ2A* (intronic) | 12:57023283 | G/C | European | 55653 | NA | 2.75 (0.43) | 1.74E-10 | 0.22 | 0.22 | 3.01 (0.96) | 1.80E-03 | 5.81E-03 | 2.79 (0.39) | 1.24E-12 | [1] |
| rs3184504 | *SH2B3* (nonsynonymous) | 12:111884607 | T/C | European | 56354 | NA | 3.99 (0.37) | 1.22E-26 | 0.27 | 0.27 | 3.75 (0.96) | 8.61E-05 | 5.17E-04 | 3.96 (0.35) | 6.19E-30 | [1] |
| rs739496 | *SH2B3* (3' UTR) | 12:111887659 | A/G | Japanese | 14806 | 0.16 | -0.14 (0.02) | 4.75E-19 | 0.31 | 0.69 | -0.06 (0.01) | 2.94E-05 | 3.61E-05 | -0.09 (0.01) | 3.59E-19 | [6] |
| rs11065987 | *BRAP* (intergenic) | 12:112072424 | G/A | European | 9316 | 0.34 | 5.07 (0.69) | 2.20E-13 | 0.25 | 0.25 | 3.65 (0.97) | 1.79E-04 | 1.79E-04 | 4.60 (0.56) | 3.75E-16 | [4] |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs6490294 | *ACAD10* (intronic) | 12:112190438 | C/A | African | 16388 | 0.34 | -4.38 (0.75) | 4.78E-09 | 0.37 | 0.63 | -2.5 (0.86) | 3.44E-03 | 4.91E-03 | -3.56 (0.56) | 2.62E-10 | [5] |
| rs4148441 | *ABCC4* (intronic) | 13:95898206 | G/A | European | 64120 | NA | 4.12 (0.6) | 6.76E-12 | 0.14 | 0.86 | 4.19 (1.14) | 2.48E-04 | 1.16E-03 | 4.13 (0.53) | 7.32E-15 | [1] |
| rs11628318 | *ANKRD9-RCOR1* (intergenic) | 14:103040086 | A/T | European | 62438 | NA | 2.57 (0.41) | 2.04E-10 | 0.38 | 0.62 | 2.07 (0.83) | 1.27E-02 | 2.96E-02 | 2.48 (0.36) | 1.04E-11 | [1] |
| rs2297067 | *EXOC3L4* (nonsynonymous) | 14:103566784 | T/C | European | 41687 | NA | 3.54 (0.55) | 1.58E-10 | 0.20 | 0.20 | 2.65 (0.98) | 7.04E-03 | 1.83E-02 | 3.32 (0.48) | 5.29E-12 | [1] |
| rs10512472 | *SLFN14* (nonsynonymous) | 17:33884803 | C/T | European | 58692 | NA | 3.64 (0.48) | 2.40E-14 | 0.23 | 0.23 | 3.08 (0.94) | 1.10E-03 | 4.63E-03 | 3.52 (0.43) | 1.29E-16 | [1] |
| rs11082304 | *CABLES1* (intronic) | 18:20720972 | G/T | European | 58215 | NA | 2.48 (0.38) | 5.27E-11 | 0.36 | 0.64 | 3.36 (0.83) | 5.01E-05 | 3.50E-04 | 2.63 (0.34) | 1.97E-14 | [1] |
| rs8109288 | *TPM4* (intronic) | 19:16185558 | G/A | European | 29014 | NA | 11.95 (1.89) | 2.75E-10 | 0.03 | 0.97 | 13.83 (2.46) | 2.01E-08 | 3.73E-05 | 12.64 (1.5) | 3.61E-17 | [1] |
| rs8109288 | *TPM4* (intronic) | 19:16185559 | A/G | African | 16388 | 0.10 | -8.72 (1.4) | 5.02E-10 | 0.03 | 0.03 | -13.83 (2.46) | 2.01E-08 | 9.81E-08 | -9.97 (1.22) | 2.67E-16 | [5] |
| rs17356664 | *EXCOC3L2-MARK4* (intergenic) | 19:45740770 | C/T | European | 55487 | NA | 2.60 (0.42) | 3.60E-10 | 0.19 | 0.81 | 2.81 (1.05) | 7.41E-03 | 1.83E-02 | 2.63 (0.39) | 9.86E-12 | [1] |
| rs2336384 | *MFN2* (intronic) | 1:12046062 | G/T | European | 57366 | NA | 2.17 (0.38) | 1.25E-08 | 0.43 | 0.43 | 1.12 (0.81) | 1.67E-01 | 2.42E-01 | 1.98 (0.35) | 9.99E-09 | [1] |
| rs10914144 | *DNM3* (intronic) | 1:171949749 | T/C | European | 54978 | NA | 3.42 (0.49) | 2.22E-12 | 0.22 | 0.22 | 1.94 (0.99) | 4.96E-02 | 8.68E-02 | 3.13 (0.44) | 7.97E-13 | [1] |
| rs1668871 | *TMCC2* (intronic) | 1:205237136 | C/T | European | 58108 | NA | 2.80 (0.37) | 2.59E-14 | 0.23 | 0.23 | -0.41 (0.94) | 6.62E-01 | 1.00E+00 | 2.38 (0.34) | 4.09E-12 | [1] |
| rs7550918 | *GCSAMLL* (intronic) | 1:247675558 | T/C | European | 54171 | NA | 3.13 (0.47) | 2.91E-11 | 0.27 | 0.73 | 0.97 (0.9) | 2.78E-01 | 3.54E-01 | 2.67 (0.42) | 1.62E-10 | [1] |
| rs17030845 | *THADA* (intronic) | 2:43687878 | C/T | European | 65738 | NA | 3.58 (0.56) | 1.27E-10 | 0.08 | 0.92 | 1.13 (1.49) | 4.50E-01 | 4.97E-01 | 3.28 (0.52) | 3.14E-10 | [1] |
| rs7694379 | *KLHL8-HSD17B13* (intergenic) | 4:88186508 | A/G | European | 56430 | NA | 2.13 (0.37) | 8.70E-09 | 0.32 | 0.32 | 0.02 (0.87) | 9.83E-01 | 1.00E+00 | 1.80 (0.34) | 1.16E-07 | [1] |
| rs17568628 | *F2R-F2RL1* (intergenic) | 5:76046938 | T/C | European | 44759 | NA | 6.07 (0.99) | 9.61E-10 | 0.02 | 0.98 | 6.13 (3.02) | 4.24E-02 | 7.74E-02 | 6.08 (0.94) | 1.16E-10 | [1] |
| rs2070729 | *IRF1* (intronic) | 5:131819920 | A/C | European | 56469 | NA | 2.39 (0.37) | 1.13E-10 | 0.45 | 0.55 | 0.58 (0.82) | 4.76E-01 | 5.12E-01 | 2.08 (0.34) | 6.85E-10 | [1] |
| rs13236689 | *CD36* (intronic) | 7:80236014 | G/T | African | 16388 | 0.44 | 4.18 (0.7) | 2.84E-09 | 0.36 | 0.36 | 1.64 (0.84) | 5.00E-02 | 6.02E-02 | 3.14 (0.54) | 5.26E-09 | [5] |
| rs4731120 | WASL-HYALP1 (intergenic) | 7:123411222 | C/A | European | 66147 | NA | 4.14 (0.59) | 2.77E-12 | 0.06 | 0.06 | 2.25 (1.63) | 1.67E-01 | 2.42E-01 | 3.92 (0.56) | 1.85E-12 | [1] |
| rs4246215 | *FEN1* (3' UTR) | 11:61564298 | T/G | European | 56299 | NA | 2.45 (0.39) | 3.31E-10 | 0.49 | 0.49 | 1.87 (0.9) | 3.85E-02 | 7.35E-02 | 2.36 (0.36) | 4.39E-11 | [1] |
| rs477895 | *BAD* (intronic) | 11:64048912 | C/T | African | 16388 | 0.45 | -4.19 (0.77) | 4.91E-08 | 0.21 | 0.21 | -1.94 (1.01) | 5.42E-02 | 6.02E-02 | -3.36 (0.61) | 3.99E-08 | [5] |
| rs17824620 | *RPH3A* (intronic) | 12:113100993 | C/A | European | 51530 | NA | 2.46 (0.43) | 9.67E-09 | 0.32 | 0.68 | 1.09 (0.86) | 2.10E-01 | 2.93E-01 | 2.19 (0.38) | 1.19E-08 | [1] |
| rs7961894 | *WDR66* (intronic) | 12:122365582 | C/T | European | 51897 | NA | 3.92 (0.61) | 1.22E-10 | 0.07 | 0.93 | 1.97 (1.6) | 2.18E-01 | 2.95E-01 | 3.68 (0.57) | 1.06E-10 | [1] |
| rs8022206 | *RAD61B* (intronic) | 14:68520905 | G/A | European | 52251 | NA | 3.20 (0.5) | 1.55E-10 | 0.18 | 0.82 | 0.49 (1.03) | 6.36E-01 | 6.67E-01 | 2.68 (0.45) | 2.51E-09 | [1] |
| rs8006385 | *ITPK1* (intronic) | 14:93501025 | G/A | European | 64929 | NA | 3.59 (0.56) | 1.24E-10 | 0.11 | 0.11 | 1.12 (1.29) | 3.85E-01 | 4.49E-01 | 3.2 (0.51) | 4.26E-10 | [1] |
| rs7149242 | *BEGAIN-DLK1* (intergenic) | 14:101159415 | G/T | European | 61247 | NA | 2.14 (0.39) | 2.68E-08 | 0.32 | 0.68 | 0.73 (0.87) | 4.02E-01 | 4.56E-01 | 1.91 (0.35) | 5.77E-08 | [1] |
| rs3809566 | *TPM1* (5' upstream) | 15:63333723 | G/A | European | 57113 | NA | 2.44 (0.39) | 3.65E-10 | 0.27 | 0.73 | 1.08 (0.89) | 2.26E-01 | 2.96E-01 | 2.22 (0.36) | 4.84E-10 | [1] |
| rs1719271 | *PLEKHO2-ANKDD1A* (intergenic) | 15:65183800 | G/A | European | 56782 | NA | 3.41 (0.5) | 1.05E-11 | 0.27 | 0.27 | 0.94 (0.92) | 3.04E-01 | 3.76E-01 | 2.85 (0.44) | 1.05E-10 | [1] |
| rs6065 | *GP1BA* (nonsynonymous) | 17:4836380 | T/C | European | 64987 | NA | 4.19 (0.63) | 2.92E-11 | 0.14 | 0.14 | 2.13 (1.15) | 6.34E-02 | 1.07E-01 | 3.71 (0.55) | 1.77E-11 | [1] |
| rs397969 | *ULK2-AKAP10* (intergenic) | 17:19804246 | C/T | European | 60944 | NA | 2.13 (0.36) | 2.32E-09 | 0.43 | 0.43 | 0.71 (0.8) | 3.78E-01 | 4.49E-01 | 1.89 (0.33) | 6.23E-09 | [1] |
| rs559972 | *TAOK1* (intronic) | 17:27814495 | T/C | European | 53460 | NA | 3.26 (0.38) | 3.30E-18 | 0.39 | 0.39 | 1.19 (0.81) | 1.43E-01 | 2.22E-01 | 2.9 (0.34) | 1.64E-17 | [1] |
| rs708382 | *FAM171A2* (5' upstream) | 17:42442343 | T/C | European | 50036 | NA | 2.44 (0.43) | 1.51E-08 | 0.37 | 0.63 | 1.81 (0.81) | 2.55E-02 | 5.10E-02 | 2.3 (0.38) | 1.50E-09 | [1] |
| rs151361 | *SLMO2* (intronic) | 20:57614002 | G/A | African | 16388 | 0.26 | 4.49 (0.78) | 9.44E-09 | 0.21 | 0.21 | 1.51 (0.98) | 1.22E-01 | 1.22E-01 | 3.33 (0.61) | 4.73E-08 | [5] |
| rs1034566 | *ARVCF* (intronic) | 22:19984276 | T/C | European | 61469 | NA | 2.13 (0.38) | 3.06E-08 | 0.21 | 0.21 | 1.79 (0.98) | 6.87E-02 | 1.11E-01 | 2.08 (0.36) | 5.73E-09 | [1] |

SNPs previously reported as genome-wide significant in GWA studies of platelet count that met or exceeded the generalization p-value are included. [a]Allele frequencies reported for 1000G CEU, not study population, by Soranzo, et al, 2009.,MAF = Minor Allele Frequency; CAF = Coded Allele Frequency; SE= Standard Error. Study population allele frequency not reported (NA) by Gieger, et al, 2011.

**Table S4**. Count of European, African, and Native American ancestral alleles vs. genotype rs117672662 on chromosome 14

| | **Ancestral Allele Count** | **rs117672662 Genotype** | | |
|---|---|---|---|---|
| | | **0** | **1** | **2** |
| European | **0** | 64 | 743 | 2,325 |
| | **1** | 0 | 625 | 4,797 |
| | **2** | 0 | 16 | 4,223 |
| African | **0** | 64 | 1,246 | 8,450 |
| | **1** | 0 | 138 | 2,338 |
| | **2** | 0 | 0 | 557 |
| Native American | **0** | 0 | 19 | 6,625 |
| | **1** | 0 | 757 | 3,445 |
| | **2** | 64 | 608 | 1,275 |

**Table S5.** Association of novel PLT variants from the HCHS/SOL discovery analysis with mean platelet volume (MPV) in a subset of the replication samples (N=4041)

| Annotated Gene(s) (location) | Chromosome: Position (build 37-hg19) | rsID | Coded/ Alt. Allele on (+) strand | WHI | | | Bio*Me* | | | Meta-analysis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | N | Beta (SE) | p-value | N | Beta (SE) | p-value | N | Beta (SE) | p-value |
| *ACTN1* (intronic) | chr14:69425467 | rs117672662 | T/C | 1,187 | 0.46 (0.09) | 6.5E-08 | 2,854 | 0.63 (0.09) | 6.7E-12 | 4,041 | 0.54 (0.06) | 3.9E-18 |
| *ZBTB9-BAK1* (intergenic) | chr6:33524820 | rs62405954 | T/C | 1,187 | -0.11 (0.08) | 0.156 | 2,854 | -0.10 (0.06) | 0.112 | 4,041 | -0.10 (0.05) | 0.032 |
| *GABBR1-MOG* (intergenic) | chr6:29608184 | rs75140056 | C/CAT | 1,187 | 0.15 (0.06) | 0.011 | 2,854 | 0.06 (0.03) | 0.049 | 4,041 | 0.08 (0.03) | 0.003 |
| *ETV7* (intronic) | chr6:36344980 | rs9470264 | G/A | 1,187 | -0.16 (0.03) | 0.009 | 2,854 | -0.24 (0.05) | 4.5E-07 | 4,041 | -0.18 (0.02) | 1.9E-15 |
| *MEF2C* (intronic) | chr5:88133921 | rs144261491 | C/T | 1,187 | 0.12 (0.15) | 0.415 | NA | NA | NA | 4,041 | 0.12 (0.15) | 0.42 |

SE= Standard Error

**Table S6.** Datasets used for functional annotation and their sources.

| Dataset[a] | Project/Study | Download/Access resource |
|---|---|---|
| All datasets of CD34-negative, CD41-positive, CD42-positive megakaryocyte cells | BLUEPRINT[7] | UCSC browser, BLUEPRINT trackhub |
| All K562 datasets | ENCODE[8] | UCSC browser, ENCODE Analysis trackhub and Integrated Regulation from ENCODE Tracks |
| GSM607949(GATA1 in megakaryocytes) | Tjissen et al, 2011[9] | CODEX,http://codex.stemcells.cam.ac.uk/ |
| GSM607950(GATA2 in megakaryocytes) | Tjissen et al, 2011[9] | CODEX,http://codex.stemcells.cam.ac.uk/ |
| GSM607951(RUNX1 in megakaryocytes) | Tjissen et al, 2011[9] | CODEX,http://codex.stemcells.cam.ac.uk/ |
| GSM607952(FLI1 in megakaryocytes) | Tjissen et al, 2011[9] | CODEX,http://codex.stemcells.cam.ac.uk/ |
| GSM607953(TAL1 in megakaryocytes) | Tjissen et al, 2011[9] | CODEX,http://codex.stemcells.cam.ac.uk/ |
| GSE23730(SKNO-1_ERG) | Martens et al,2012[10] | BLOODCHIP, http://149.171.101.136/python/BloodChIP/ |
| GSE23730(SKNO-1_FLI1) | Martens et al,2012[10] | BLOODCHIP, http://149.171.101.136/python/BloodChIP/ |
| GSE23730(SKNO-1_RUNX1) | Martens et al,2012[10] | BLOODCHIP, http://149.171.101.136/python/BloodChIP/ |

[a]All datasets were mapped against Human Genome Build 37/hg19

**Table S7.** Significant Blood Browser[11] cis-eQTL for Results for PLT-associated SNPs and their LD partners ($r^2 \geq 0.4$)

| Gene | Index rsID | Proxy rsID | Chr: Position (build 37/hg19) | $r^2$ | type[a] | p-value | Probe Center Position | Alleles on (+) strand | Allele Assessed | Overall Z-Score | HUGO Gene Name | FDR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *ETV7* | rs9470264 | rs7758498 | 6:36344213 | 0.53 | 0 | 3.80E-18 | 6:36686912 | G/A | A | -8.69 | - | 0 |
| *ETV7* | rs9470264 | rs7758498 | 6:36344213 | 0.53 | 0 | 6.50E-07 | 6:36570090 | G/A | A | 4.98 | *STK38* | 0 |
| *ETV7* | rs9470264 | rs7758498 | 6:36344213 | 0.53 | 0 | 1.10E-06 | 6:36565873 | G/A | A | -4.87 | *KCTD20* | 0 |
| *ETV7* | rs9470264 | rs7758498 | 6:36344213 | 0.53 | 0 | 2.10E-06 | 6:36678715 | G/A | A | -4.75 | *SFRS3* | 0.001 |
| *ETV7* | rs9470264 | rs1998266 | 6:36358289 | 0.58 | 2 | 3.10E-28 | 6:36686912 | C/T | T | -11.02 | - | 0 |
| *ETV7* | rs9470264 | rs1998266 | 6:36358289 | 0.58 | 2 | 2.70E-08 | 6:36678715 | C/T | T | -5.56 | *SFRS3* | 0 |
| *ETV7* | rs9470264 | rs1998266 | 6:36358289 | 0.58 | 2 | 1.70E-07 | 6:36570090 | C/T | T | 5.23 | *STK38* | 0 |
| *ETV7* | rs9470264 | rs4713971 | 6:36369786 | 0.40 | 0 | 2.60E-23 | 6:36686912 | G/C | G | -9.95 | - | 0 |
| *ETV7* | rs9470264 | rs4713971 | 6:36369786 | 0.40 | 0 | 3.30E-08 | 6:36678715 | G/C | G | -5.53 | *SFRS3* | 0 |
| *ETV7* | rs9470264 | rs4713971 | 6:36369786 | 0.40 | 0 | 3.10E-07 | 6:36570090 | G/C | G | 5.12 | *STK38* | 0 |
| *ETV7* | rs9470264 | rs941816 | 6:36375304 | 0.40 | 2 | 9.60E-24 | 6:36686912 | G/A | G | -10.05 | - | 0 |
| *ETV7* | rs9470264 | rs941816 | 6:36375304 | 0.40 | 2 | 4.00E-08 | 6:36678715 | G/A | G | -5.49 | *SFRS3* | 0 |
| *ETV7* | rs9470264 | rs941816 | 6:36375304 | 0.40 | 2 | 3.00E-07 | 6:36570090 | G/A | G | 5.12 | *STK38* | 0 |
| *ETV7* | rs9470264 | rs6457915 | 6:36380644 | 0.40 | 0 | 3.30E-08 | 6:36678715 | T/C | C | -5.52 | *SFRS3* | 0 |
| *ETV7* | rs9470264 | rs6457915 | 6:36380644 | 0.40 | 0 | 2.10E-23 | 6:36686912 | T/C | C | -9.97 | - | 0 |
| *ETV7* | rs9470264 | rs6457915 | 6:36380644 | 0.40 | 0 | 2.50E-07 | 6:36570090 | T/C | C | 5.16 | *STK38* | 0 |
| *ETV7* | rs9470264 | rs4713975 | 6:36395761 | 0.59 | 2 | 2.10E-34 | 6:36686912 | G/A | A | -12.23 | - | 0 |
| *ETV7* | rs9470264 | rs4713975 | 6:36395761 | 0.59 | 2 | 1.50E-08 | 6:36565873 | G/A | A | -5.67 | *KCTD20* | 0 |
| *ETV7* | rs9470264 | rs4713975 | 6:36395761 | 0.59 | 2 | 1.10E-08 | 6:36570090 | G/A | A | 5.72 | *STK38* | 0 |
| *ETV7* | rs9470264 | rs4713975 | 6:36395761 | 0.59 | 2 | 4.60E-08 | 6:36678715 | G/A | A | -5.47 | *SFRS3* | 0 |
| *ETV7* | rs9470264 | rs1744653 | 6:36397688 | 0.46 | 0 | 6.50E-11 | 6:36678715 | G/A | A | -6.53 | *SFRS3* | 0 |
| *ETV7* | rs9470264 | rs1744653 | 6:36397688 | 0.46 | 0 | 8.00E-24 | 6:36686912 | G/A | A | -10.06 | - | 0 |
| *ETV7* | rs9470264 | rs1744653 | 6:36397688 | 0.46 | 0 | 2.40E-05 | 6:36570090 | G/A | A | 4.23 | *STK38* | 0.011 |
| *ETV7* | rs9470264 | rs7741260 | 6:36399120 | 0.46 | 2 | 6.90E-24 | 6:36686912 | C/T | T | -10.08 | - | 0 |
| *ETV7* | rs9470264 | rs7741260 | 6:36399120 | 0.46 | 2 | 5.40E-11 | 6:36678715 | C/T | T | -6.56 | *SFRS3* | 0 |
| *ETV7* | rs9470264 | rs7741260 | 6:36399120 | 0.46 | 2 | 2.40E-05 | 6:36570090 | C/T | T | 4.23 | *STK38* | 0.011 |
| *ETV7* | rs9470264 | rs16888605 | 6:36433408 | 0.57 | 0 | 3.50E-09 | 6:36565873 | A/G | G | -5.9 | *KCTD20* | 0 |
| *ETV7* | rs9470264 | rs16888605 | 6:36433408 | 0.57 | 0 | 1.40E-35 | 6:36686912 | A/G | G | -12.45 | - | 0 |
| *ETV7* | rs9470264 | rs16888605 | 6:36433408 | 0.57 | 0 | 9.40E-09 | 6:36678715 | A/G | G | -5.74 | *SFRS3* | 0 |
| *ETV7* | rs9470264 | rs16888605 | 6:36433408 | 0.57 | 0 | 2.80E-09 | 6:36570090 | A/G | G | 5.94 | *STK38* | 0 |
| *ETV7* | rs9470264 | rs4713978 | 6:36466787 | 0.56 | 2 | 1.20E-08 | 6:36678715 | C/A | A | -5.7 | *SFRS3* | 0 |
| *ETV7* | rs9470264 | rs4713978 | 6:36466787 | 0.56 | 2 | 1.40E-09 | 6:36570090 | C/A | A | 6.06 | *STK38* | 0 |
| *ETV7* | rs9470264 | rs4713978 | 6:36466787 | 0.56 | 2 | 1.60E-36 | 6:36686912 | C/A | A | -12.62 | - | 0 |
| *ETV7* | rs9470264 | rs4713978 | 6:36466787 | 0.56 | 2 | 7.40E-10 | 6:36565873 | C/A | A | -6.16 | *KCTD20* | 0 |
| *GABBR1-MOG* | rs75140056 | rs29273 | 6:29610989 | 0.45 | 2 | 2.10E-09 | 6:29811988 | G/C | C | 5.99 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs29273 | 6:29610989 | 0.45 | 2 | 3.40E-21 | 6:29906646 | G/C | C | 9.45 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs29273 | 6:29610989 | 0.45 | 2 | 1.10E-48 | 6:29800135 | G/C | C | -14.66 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs9257928 | 6:29613218 | 0.42 | 0 | 3.20E-35 | 6:29800135 | C/T | T | -12.38 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs9257928 | 6:29613218 | 0.42 | 0 | 4.60E-10 | 6:29811988 | C/T | T | 6.23 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs9257928 | 6:29613218 | 0.42 | 0 | 1.40E-09 | 6:29906646 | C/T | T | 6.06 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs29269 | 6:29617747 | 0.45 | 2 | 1.10E-48 | 6:29800135 | C/T | T | -14.67 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs29269 | 6:29617747 | 0.45 | 2 | 1.60E-09 | 6:29811988 | C/T | T | 6.04 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs29269 | 6:29617747 | 0.45 | 2 | 3.70E-21 | 6:29906646 | C/T | T | 9.44 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs29231 | 6:29618525 | 0.44 | 2 | 2.40E-09 | 6:29811988 | C/T | T | 5.97 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs29231 | 6:29618525 | 0.44 | 2 | 1.10E-46 | 6:29800135 | C/T | T | -14.35 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs29231 | 6:29618525 | 0.44 | 2 | 1.80E-20 | 6:29906646 | C/T | T | 9.28 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs2535246 | 6:29636409 | 0.42 | 2 | 3.40E-10 | 6:29811988 | T/G | G | 6.28 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs2535246 | 6:29636409 | 0.42 | 2 | 1.20E-20 | 6:29906646 | T/G | G | 9.32 | *HLA-G* | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *GABBR1-MOG* | rs75140056 | rs2535246 | 6:29636409 | 0.42 | 2 | 3.40E-47 | 6:29800135 | T/G | G | -14.43 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs2535246 | 6:29636409 | 0.42 | 2 | 3.60E-05 | 6:29783657 | T/G | G | 4.13 | - | 0.016 |
| *GABBR1-MOG* | rs75140056 | rs1122947 | 6:29638434 | 0.42 | 2 | 2.70E-47 | 6:29800135 | A/G | G | -14.44 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs1122947 | 6:29638434 | 0.42 | 2 | 3.40E-10 | 6:29811988 | A/G | G | 6.28 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs1122947 | 6:29638434 | 0.42 | 2 | 1.00E-20 | 6:29906646 | A/G | G | 9.33 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs1122947 | 6:29638434 | 0.42 | 2 | 3.60E-05 | 6:29783657 | A/G | G | 4.13 | - | 0.016 |
| *GABBR1-MOG* | rs75140056 | rs2747453 | 6:29654943 | 0.44 | 2 | 1.40E-49 | 6:29800135 | T/C | C | -14.8 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs2747453 | 6:29654943 | 0.44 | 2 | 1.40E-22 | 6:29906646 | T/C | C | 9.78 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs2747453 | 6:29654943 | 0.44 | 2 | 1.20E-08 | 6:29811988 | T/C | C | 5.7 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs3129090 | 6:29664131 | 0.44 | 2 | 7.60E-44 | 6:29800135 | C/T | T | -13.89 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs3129090 | 6:29664131 | 0.44 | 2 | 2.90E-09 | 6:29811988 | C/T | T | 5.94 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs3129090 | 6:29664131 | 0.44 | 2 | 8.50E-22 | 6:29906646 | C/T | T | 9.59 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs3131879 | 6:29665458 | 0.43 | 0 | 2.40E-09 | 6:29811988 | A/C | C | 5.97 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs3131879 | 6:29665458 | 0.43 | 0 | 1.30E-22 | 6:29906646 | A/C | C | 9.78 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs3131879 | 6:29665458 | 0.43 | 0 | 1.00E-41 | 6:29800135 | A/C | C | -13.53 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs9258102 | 6:29668913 | 0.43 | 2 | 5.40E-22 | 6:29906646 | T/C | C | 9.64 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs9258102 | 6:29668913 | 0.43 | 2 | 4.70E-44 | 6:29800135 | T/C | C | -13.92 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs9258102 | 6:29668913 | 0.43 | 2 | 2.80E-09 | 6:29811988 | T/C | C | 5.94 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs9258114 | 6:29669640 | 0.43 | 2 | 7.10E-09 | 6:29811988 | C/A | A | 5.79 | *AL645939.6-3* | 0 |
| *GABBR1-MOG* | rs75140056 | rs9258114 | 6:29669640 | 0.43 | 2 | 1.10E-62 | 6:29800135 | C/A | A | -16.71 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs9258114 | 6:29669640 | 0.43 | 2 | 8.60E-19 | 6:29906646 | C/A | A | 8.85 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs6456993 | 6:29670536 | 0.43 | 0 | 1.70E-21 | 6:29906646 | G/A | A | 9.52 | *HLA-G* | 0 |
| *GABBR1-MOG* | rs75140056 | rs6456993 | 6:29670536 | 0.43 | 0 | 7.20E-45 | 6:29800135 | G/A | A | -14.05 | *HLA-F* | 0 |
| *GABBR1-MOG* | rs75140056 | rs6456993 | 6:29670536 | 0.43 | 0 | 8.60E-10 | 6:29811988 | G/A | A | 6.13 | *AL645939.6-3* | 0 |
| *BAK1* | rs62405954 | rs12206050 | 6:33564296 | 0.46 | 0 | 3.18E-05 | 6:33493262 | A/T | T | 4.16 | *CUTA* | 0.014 |
| *BAK1* | rs62405954 | rs12206050 | 6:33564296 | 0.46 | 0 | 3.88E-12 | 6:33492459 | A/T | T | 6.94 | *CUTA* | 0 |
| *BAK1* | rs62405954 | rs12664430 | 6:33272677 | 0.62 | 2 | 4.59E-09 | 6:33162754 | C/T | T | -5.86 | *HLA-DPB1* | 0 |
| *BAK1* | rs62405954 | rs12664430 | 6:33272677 | 0.62 | 2 | 2.19E-04 | 6:33533115 | C/T | T | 3.7 | *SYNGAP1,ZBTB9* | 0.078 |
| *BAK1* | rs62405954 | rs12664430 | 6:33272677 | 0.62 | 2 | 3.88E-08 | 6:33354286 | C/T | T | -5.5 | *B3GALT4* | 0 |
| *BAK1* | rs62405954 | rs12664430 | 6:33272677 | 0.62 | 2 | 1.02E-06 | 6:33375735 | C/T | T | 4.89 | *TAPBP* | 3.02E-04 |
| *BAK1* | rs62405954 | rs12664430 | 6:33272677 | 0.62 | 2 | 3.67E-04 | 6:33379629 | C/T | T | 3.56 | *TAPBP* | 0.117 |
| *BAK1* | rs62405954 | rs12664430 | 6:33272677 | 0.62 | 2 | 3.47E-03 | 6:33492459 | C/T | T | 2.92 | *CUTA* | 0.496 |
| *BAK1* | rs62405954 | rs3846855 | 6:33555877 | 0.47 | 0 | 1.48E-10 | 6:33492459 | G/A | A | 6.41 | *CUTA* | 0 |
| *BAK1* | rs62405954 | rs3846855 | 6:33555877 | 0.47 | 0 | 1.61E-05 | 6:33493262 | G/A | A | 4.31 | *CUTA* | 0.007 |
| *BAK1* | rs62405954 | rs5745568 | 6:33548394 | 0.46 | 2 | 1.89E-08 | 6:33492459 | G/T | T | 5.62 | *CUTA* | 0 |
| *BAK1* | rs62405954 | rs5745568 | 6:33548394 | 0.46 | 2 | 5.95E-05 | 6:33493262 | G/T | T | 4.01 | *CUTA* | 0.025 |
| *BAK1* | rs62405954 | rs9380365 | 6:33472317 | 0.42 | 2 | 5.06E-10 | 6:33354286 | T/A | A | -6.22 | *B3GALT4* | 0 |
| *BAK1* | rs62405954 | rs9380365 | 6:33472317 | 0.42 | 2 | 1.95E-07 | 6:33491946 | T/A | A | 5.2 | *PHF1* | 6.31E-05 |
| *BAK1* | rs62405954 | rs9380365 | 6:33472317 | 0.42 | 2 | 7.90E-06 | 6:33533115 | T/A | A | 4.47 | *SYNGAP1,ZBTB9* | 0.004 |
| *BAK1* | rs62405954 | rs9380365 | 6:33472317 | 0.42 | 2 | 7.41E-04 | 6:33493262 | T/A | A | 3.37 | *CUTA* | 0.198 |
| *BAK1* | rs62405954 | rs9380365 | 6:33472317 | 0.42 | 2 | 5.22E-17 | 6:33375735 | T/A | A | 8.38 | *TAPBP* | 0 |
| *BAK1* | rs62405954 | rs9380365 | 6:33472317 | 0.42 | 2 | 1.95E-07 | 6:33491946 | T/A | A | 5.2 | *PHF1* | 6.31E-05 |
| *BAK1* | rs62405954 | rs9380365 | 6:33472317 | 0.42 | 2 | 7.90E-06 | 6:33533115 | T/A | A | 4.47 | *SYNGAP1,ZBTB9* | 0.004 |
| *BAK1* | rs62405954 | rs9380365 | 6:33472317 | 0.42 | 2 | 7.41E-04 | 6:33493262 | T/A | A | 3.37 | *CUTA* | 0.198 |
| *BAK1* | rs62405954 | rs9380365 | 6:33472317 | 0.42 | 2 | 5.22E-17 | 6:33375735 | T/A | A | 8.38 | *TAPBP* | 0 |

[a]Type: Imputed=0, genotyped=2

**Table S8.** Summary of findings from the functional annotation of the 5 novel SNPs and LD partners ($r^2 \geq 0.5$) identified in HCHS/SOL

| Index SNP rsID | fSNP* rsID | $r^2$ | fSNP p-value | fSNP Beta (SE) | Nearest Genes to fSNP | coded/ alternate fSNP allele on (+) strand | fSNP CAF | fSNP oevar[a] | 1000 genomes phase III coded allele frequency | | | | | | Description of Findings |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | African | European | Ad Mixed American | South Asian | East Asian | All | |
| rs75140056 | rs29269 | 0.5 | 1.23E-06 | 0.131 (0.029) | *GABBR1* | C/T | 0.77 | 1.0 | 0.86 | 0.81 | 0.8 | 0.86 | 0.95 | 0.86 | rs29269 lies in a putative active *GABBR1* promoter in megakaryocytes, also overlaps several transcription factor ChIP-Seq peaks including TBP and ETS1 in K562 cells. Also overlaps RNA PolI ChIP-Seq Peak |
| rs62405954 | rs1002011 | 0.6 | 7.92E-06 | -0.186 (0.044) | *VPS52* | G/A | 0.89 | 1.0 | 1.00 | 1.00 | 0.88 | 1.00 | 1.00 | 0.98 | Proxy SNP rs1002011 lies within a putative enhancer overlapping the 5' UTR of the *VPS52* gene. It also lies within a DNase1 hypersensitive peak and CFOS ChIP-Seq peak in HUVEC cells. |
| rs9470264 | rs80331350 | 0.8 | 2.66E-05 | -0.144 (0.034) | *PXT1* | T/C | 0.81 | 1.1 | 0.87 | 0.99 | 0.78 | 0.99 | 0.87 | 0.9 | rs80331350 lies in a putative megakaryocyte enhancer in *PXT1* intron. The putative enhancer element overlaps GATA1 ChIP-Seq peak in peripheral blood-derived erythroblasts (PBDE) cells |
| rs9470264 | rs2273883 | 0.8 | 2.18E-05 | -0.170 (0.035) | *KCTD20* | C/A | 0.82 | 1.1 | 0.99 | 0.99 | 0.79 | 0.99 | 0.87 | 0.94 | rs227883 (r2=0.8) lies within RUNX1 bound putative active *KCTD20* promoter in megakaryocytes. The putative promoter element overlaps ChIP-Seq peaks of promoter binding proteins RNA polymerase II, TBP, P300 and key megakaryocyte regulators including GATA1, GATA2, ETS1 and GABPA in K562 cells |
| rs9470264 | rs113553570 | 0.8 | 1.05E-06 | -0.168 (0.035) | *STK38* | G/A | 0.82 | 1.1 | 0.89 | 0.99 | 0.78 | 0.99 | 0.87 | 0.91 | proxy SNP rs113553570 lies in a putative enhancer in STK38 intron |
| rs9470264 | rs2239541 | 0.8 | 1.11E-06 | -0.168 (0.035) | *STK38* | G/A | 0.82 | 1.1 | 0.89 | 0.99 | 0.78 | 0.99 | 0.87 | 0.91 | proxy SNP rs2239541 (r2=0.8), which lies in a putative megakaryocyte enhancer in STK38 intron |
| rs9470264 | rs201891379 | 0.8 | 1.11E-06 | -0.172 (0.034) | *STK38* | ACAATACAATTAACTAAAC/A | 0.81 | 1.1 | 0.87 | 0.99 | 0.8 | 0.99 | 0.87 | 0.91 | proxy indel rs201891379 (r2=0.8), which lies in a putative promoter element upstream of STK38 gene. The putative enhancer element overlaps ChIP-Seq peaks of GABP and IRF1 in K562 cells |
| rs117672662 | rs117672662 | 1.0 | 4.88E-07 | 0.609 (0.055) | *ACTN1* | T/C | 0.94 | 1.0 | 1.00 | 1.00 | 0.93 | 1.00 | 1.00 | 0.99 | rs117672662 lies in a megakaryocyte-specific putative enhancer containing a RUNX1 consensus sequence and overlaps ChIP-Seq peaks of key megakaryocyte regulators ERG, FLI1 and RUNX1 in SKNO-1 cells (an acute myeloid leukemia cell line) |
| rs144261491 | rs200572016 | 0.8 | 8.65E-29 | 0.393 (0.081) | *MEF2C-AS1* | TGA/T | 0.97 | 0.9 | 1.00 | 1.00 | 0.96 | 1.00 | 1.00 | 0.99 | indel rs200572016 lies in a putative megakaryocyte-specific enhancer in *MEF2C-AS1* intron; this enhancer element is DNAse accessible in K562 cells and overlaps ChIP-Seq peaks of GATA2, TAL1 and P300 in these cells |

Abbreviations: fSNP = Predicted functional single nucleotide polymorphism; [a] = ratio of the variances of the observed and the estimated allele counts; CAF= Coded Allele Frequency

**Table S9**. Summary of *in silico* functional prediction algorithm results for significant variants and their LD partners

| Locus | rsID | Chr:Position | $r^2$ with HCHS/SOL lead SNP | CADD[12] C-Score | GWAVA[13] Score | RegulomeDB[14] Score | deltaSVM[1][5] Score for K562 cell line | Prioritized functional variant |
|---|---|---|---|---|---|---|---|---|
| *ACTN1* | rs117672662 | 14:69425467 | Lead | 0.56 | 0.32 | 4 | 9.84 | fSNP |
| *ACTN1* | rs140754108 | 14:69407222 | 0.84 | 6.53 | 0.26 | 2b | 0.69 | - |
| *ACTN1* | rs12431622 | 14:69415032 | 0.81 | 2.93 | 0.32 | 5 | -1.45 | - |
| *GABBR1-MOG* | rs75140056 (indel) | 6:29608184 | Lead | 7.06 | NA | NA | NA | - |
| *GABBR1-MOG* | rs3131857 | 6:29606794 | > 0.99 | 1.38 | 0.2 | 6 | -0.46 | - |
| *GABBR1-MOG* | rs4713235 | 6:29670478 | 0.60 | 4.29 | 0.34 | 2a | 1.13 | - |
| *GABBR1-MOG* | rs3131875 | 6:29666111 | 0.50 | 16.75 | 0.09 | 5 | -2.03 | - |
| *GABBR1-MOG* | NA (indel) | 6:29608870 | 0.50 | NA | NA | NA | NA | - |
| *GABBR1-MOG* | rs29269 | 6:29617747 | 0.50 | 7.00 | 0.71 | 4 | -4.58 | fSNP |
| *ETV7* | rs9470264 | 6:36344980 | Lead | 0.13 | 0.14 | 6 | 2.04 | - |
| *ETV7* | rs3778028 | 6:36349666 | 0.94 | 5.87 | 0.37 | 3a | -1.62 | - |
| *ETV7* | rs1885206 | 6:36351095 | 0.93 | 3.35 | 0.42 | 2b | -0.37 | - |
| *ETV7* | rs76676074 | 6:36361063 | 0.90 | 8.84 | 0.21 | 5 | 0.35 | - |
| *ETV7* | rs111654791 | 6:36376021 | 0.91 | 0.51 | NA | No Data | -0.15 | - |
| *ETV7* | rs80047437 | 6:36376369 | 0.91 | 4.63 | 0.11 | 6 | 0.23 | - |
| *ETV7* | rs112394821 | 6:36379855 | 0.91 | 0.03 | 0.02 | No Data | -1.88 | - |
| *ETV7* | rs113849189 | 6:36383209 | 0.91 | 2.27 | 0.03 | No Data | -0.56 | - |
| *ETV7* | rs79235676 | 6:36388019 | 0.91 | 8.38 | 0.1 | 6 | -1.55 | - |
| *ETV7* | rs74555455 | 6:36389418 | 0.91 | 3.58 | 0.1 | 6 | -2.95 | - |
| *ETV7* | rs80331350 | 6:36391236 | 0.91 | 13.53 | 0.07 | 2b | 1.07 | fSNP |
| *ETV7* | rs150014990 | 6:36392403 | 0.90 | 0.72 | 0.31 | No Data | 1.29 | - |
| *ETV7* | rs200860324 | 6:36403871 | 0.87 | 3.22 | NA | 6 | -3.31 | - |
| *ETV7* | rs12526415 | 6:36403874 | 0.84 | 0.33 | 0.11 | 6 | 1.75 | - |
| *ETV7* | rs113585688 | 6:36405876 | 0.88 | 0.67 | 0.29 | No Data | -2.42 | - |
| *ETV7* | rs4711451 | 6:36410308 | 0.84 | 2.71 | 0.82 | 4 | 4.43 | - |
| *ETV7* | rs4713976 | 6:36417410 | 0.84 | 11.93 | 0.05 | 5 | 0.64 | - |
| *ETV7* | rs4713977 | 6:36424664 | 0.81 | 2.54 | 0.08 | 5 | 2.39 | - |
| *ETV7* | rs79927326 | 6:36434643 | 0.81 | 0.81 | 0.09 | 5 | 0.07 | - |
| *ETV7* | rs2273882 | 6:36437713 | 0.81 | 0.96 | 0.16 | No Data | 3.11 | - |
| *ETV7* | rs941974 | 6:36438385 | 0.81 | 0.96 | 0.03 | 6 | -5.21 | - |
| *ETV7* | rs79328581 | 6:36454464 | 0.81 | 5.45 | 0.23 | 5 | -1.02 | - |
| *ETV7* | rs75297867 | 6:36461335 | 0.81 | 0.01 | 0.19 | 5 | 1.58 | - |
| *ETV7* | rs4236051 | 6:36469821 | 0.81 | 0.97 | 0.09 | 5 | -0.88 | - |
| *ETV7* | rs2267930 | 6:36472788 | 0.81 | 4.91 | 0.05 | 5 | -0.76 | - |
| *ETV7* | rs75144098 | 6:36475146 | 0.81 | 9.50 | 0.15 | 5 | 3.97 | - |
| *ETV7* | rs111800900 | 6:36480859 | 0.81 | 7.86 | 0.1 | No Data | 2.51 | - |
| *ETV7* | rs112845031 | 6:36486403 | 0.80 | 0.14 | 0.08 | No Data | 1.94 | - |
| *ETV7* | rs75112107 | 6:36492839 | 0.80 | 4.25 | 0.1 | 5 | 0.45 | - |
| *ETV7* | rs113553570 | 6:36493878 | 0.80 | 0.17 | 0.19 | 6 | 0.12 | fSNP |
| *ETV7* | rs223954 | 6:36494053 | 0.80 | 15.13 | 0.5 | 5 | -1.94 | fSNP |
| *ETV7* | rs2273883 | 6:88133921 | 0.76 | 10.32 | 0.94 | 1f | -0.14 | fSNP |
| *MEF2C* | rs144261491 | 5:88133921 | Lead | 2.37 | 0.18 | NA | 0.71 | - |
| *MEF2C* | rs151169192 | 5:88109756 | > 0.99 | 5.25 | 0.28 | 4 | -7.26 | - |
| *MEF2C* | rs185717407 | 5:88171283 | 0.99 | 9.77 | 0.4 | 3a | -0.18 | - |
| *MEF2C* | rs187320873 | 5:88172872 | 0.99 | 13.47 | 0.61 | 4 | 3.84 | - |
| *MEF2C* | rs184480102 | 5:87997100 | 0.75 | 3.49 | 0.22 | 6 | 0.22 | - |
| *MEF2C* | rs200572016 | 5:88220439 | 0.6 | 15.59 | NA | 3a | -0.38 | fSNP |
| *BAK1* | rs141400340 | 6:33577487 | Lead | 4.07 | 0.6 | 2b | -0.84 | fSNP |
| *BAK1* | rs62405954 | 6:33524820 | Lead | 9.19 | 0.30 | 4 | -3.11 | - |
| *BAK1* | rs186384986 | 6:33467086 | 0.91 | 3.80 | 0.01 | 6 | 0.19 | - |
| *BAK1* | rs12526020 | 6:33467084 | 0.91 | 0.53 | 0.01 | 6 | -0.28 | - |
| *BAK1* | rs12664430 | 6:33272677 | 0.62 | 12.65 | 0.45 | No Data | 1.66 | - |
| *BAK1* | rs186246149 | 6:33253646 | 0.59 | 4.12 | 0.03 | No Data | -1.62 | - |
| *BAK1* | rs1002011 | 6:33218237 | 0.56 | 3.33 | 0.94 | 4 | 0.055 | fSNP |
| *BAK1* | rs144043427 | 6:33353011 | 0.53 | NA | NA | 6 | NA | - |
| *BAK1* | rs5745564 | 6:33548696 | 0.50 | NA | NA | 5 | NA | - |
| *BAK1* | rs75080135 | 6:33552707 | 0.48 | 0.79 | 0.65 | 5 | -3.05 | - |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *BAK1* | rs3846855 | 6:33555877 | 0.47 | 2.28 | 0.13 | 6 | -1.88 | - |
| *BAK1* | rs5745568 | 6:33548394 | 0.46 | NA | 0.86 | 1b | -0.63 | - |
| *BAK1* | rs9357161 | 6:33549403 | 0.46 | 3.66 | 0.36 | No Data | 0.74 | - |
| *BAK1* | rs12206050 | 6:33564296 | 0.46 | 8.33 | 0.17 | 6 | -0.27 | - |
| *BAK1* | rs72882008 | 6:33572840 | 0.45 | 1.92 | 0.06 | 4 | 2.37 | - |
| *BAK1* | rs12214883 | 6:33513968 | 0.45 | 0.074 | 0.10 | 6 | -2.15 | - |
| *BAK1* | rs11968218 | 6:33518845 | 0.45 | 4.17 | 0.05 | 5 | -1.46 | - |
| *BAK1* | rs9380365 | 6:33472317 | 0.42 | 5.31 | 0.08 | 5 | 0 | - |
| *BAK1* | rs35144104 | 6:33463961 | 0.41 | 3.33 | 0.01 | 6 | -2.87 | - |

Abbreviation: fSNP=predicted functional single nucleotide polymorphism

**Table S10.** Results from cis-eQTL analysis of *ACTN1* rs117672662 in 1,457 Native Americans

| Chr:Position | Illumina ID | Transcript | Beta[a] (Standard Error) | P-value | SNP Number[b] |
|---|---|---|---|---|---|
| 14:66460685 | ILMN_1723607 | GPHN | -0.02 (0.08) | 8.4E-01 | 1 |
| 14:66871825 | ILMN_1750689 | MPP5 | -0.04 (0.07) | 6.0-01 | 0 |
| 14:66874720 | ILMN_1797310 | ATP6V1D | 0.04 (0.09) | 6.2E-01 | 0 |
| 14:66922877 | ILMN_1739821 | EIF2S1 | 0.12 (0.08) | 1.4E-01 | 0 |
| 14:66923813 | ILMN_1811470 | PLEK2 | -0.01 (0.07) | 8.8E-01 | 1 |
| 14:67126397 | ILMN_1798395 | PIGH | 0.10 (0.07) | 1.8E-01 | 0 |
| 14:67192886 | ILMN_1693136 | VTI1B | -0.08 (0.07) | 2.7E-01 | 1 |
| 14:67213499 | ILMN_2128741 | RDH11 | -0.07 (0.08) | 3.6E-01 | 0 |
| 14:67213871 | ILMN_1768719 | RDH11 | -0.07 (0.07) | 3.1E-01 | 0 |
| 14:67283267 | ILMN_1798061 | ZFYVE26 | 0.13 (0.07) | 6.9E-02 | 0 |
| 14:68324290 | ILMN_1675448 | ZFP36L1 | -0.01 (0.07) | 8.8E-01 | 0 |
| 14:68411163 | ILMN_2232177 | ACTN1 | -0.05 (0.09) | 5.6E-01 | 1 |
| 14:68587437 | ILMN_1735402 | WDR22 | -0.04 (0.07) | 5.7E-01 | 0 |
| 14:68778515 | ILMN_1771689 | EXD2 | 0.05 (0.07) | 5.2E-01 | 0 |
| 14:68884396 | ILMN_1799903 | GALNTL1 | -0.07 (0.07) | 3.0E-01 | 1 |
| 14:68916937 | ILMN_1781795 | ERH | 0.09 (0.07) | 2.0E-01 | 0 |
| 14:68996734 | ILMN_1756878 | SLC39A9 | 0.01 (0.08) | 9.2E-01 | 0 |
| 14:69251274 | ILMN_2226917 | KIAA0247 | 0.11 (0.07) | 1.2E-01 | 1 |
| 14:69307877 | ILMN_1761996 | SFRS5 | 0.16 (0.07) | 1.8E-02 | 0 |
| 14:69308064 | ILMN_2378868 | SFRS5 | 0.11 (0.07) | 9.4E-02 | 1 |
| 14:69862844 | ILMN_1706305 | COX16 | 0.02 (0.07) | 7.7E-01 | 0 |
| 14:69908650 | ILMN_1697793 | SYNJ2BP | -0.03 (0.08) | 6.9E-01 | 0 |
| 14:70120755 | ILMN_1654543 | MED6 | 0.06 (0.09) | 5.2E-01 | 0 |
| 14:70360380 | ILMN_1900513 | AW449499 | -0.06 (0.07) | 4.3E-01 | 0 |
| 14:70443342 | ILMN_1882112 | BX101252 | -0.14 (0.07) | 4.4E-02 | 1 |
| 14:70646545 | ILMN_1740010 | PCNX | -0.07 (0.07) | 3.4E-01 | 0 |
| 14:70934827 | ILMN_1715396 | SNORD56B | -0.13 (0.07) | 6.1E-02 | 0 |

[a] = Standard Deviation per copy of the C allele
[b]= Number of SNPs in probe

**Table S12.** Imputation scores for lead discovery variants and variants in high linkage disequilibrium ($r^2$>0.5)

| Locus_name | rsID | chromosome | position | Type[a] | $r^2$ with the lead | oevar | Info[b] |
|---|---|---|---|---|---|---|---|
| *ZBTB9-BAK1* | rs62405954 | 6 | 33524820 | 2 | 1 | 1.06 | 1.00 |
| *ZBTB9-BAK1* | rs186384986 | 6 | 33467086 | 0 | 0.9 | 1.03 | 0.97 |
| *ZBTB9-BAK1* | rs12526020 | 6 | 33467084 | 0 | 0.9 | 1.03 | 0.97 |
| *ZBTB9-BAK1* | rs12664430 | 6 | 33272677 | 2 | 0.6 | 1.04 | 1.00 |
| *ZBTB9-BAK1* | rs186246149 | 6 | 33253646 | 0 | 0.6 | 1.07 | 0.99 |
| *ZBTB9-BAK1* | rs147668544 | 6 | 33218237 | 0 | 0.6 | 0.98 | 0.90 |
| *ZBTB9-BAK1* | rs144043427 | 6 | 33353011 | 0 | 0.5 | 0.99 | 0.97 |
| *ZBTB9-BAK1* | rs5745564 | 6 | 33548696 | 0 | 0.5 | 1.05 | 1.00 |
| *ZBTB9-BAK1* | rs75080135 | 6 | 33552707 | 0 | 0.5 | 1.04 | 0.99 |
| *ZBTB9-BAK1* | rs3846855 | 6 | 33555877 | 0 | 0.5 | 1.04 | 0.99 |
| *ZBTB9-BAK1* | rs5745568 | 6 | 33548394 | 2 | 0.5 | 1.04 | 1.00 |
| *ZBTB9-BAK1* | rs9357161 | 6 | 33549403 | 0 | 0.5 | 1.03 | 1.00 |
| *ZBTB9-BAK1* | rs12206050 | 6 | 33564296 | 0 | 0.5 | 1.04 | 1.00 |
| *ZBTB9-BAK1* | rs72882008 | 6 | 33572840 | 0 | 0.5 | 1.01 | 0.98 |
| *MEF2C-AS1* | rs144261491 | 5 | 88133921 | 0 | 1 | 0.87 | 0.84 |
| *MEF2C-AS1* | rs151169192 | 5 | 88109756 | 0 | 1 | 0.87 | 0.83 |
| *MEF2C-AS1* | rs185717407 | 5 | 88171283 | 0 | 1 | 0.87 | 0.84 |
| *MEF2C-AS1* | rs187320873 | 5 | 88172872 | 0 | 1 | 0.87 | 0.84 |
| *MEF2C-AS1* | rs184480102 | 5 | 87997100 | 0 | 0.7 | 0.71 | 0.68 |
| *MEF2C-AS1* | rs200572016 | 5 | 88220439 | 0 | 0.7 | 0.89 | 0.86 |
| *MEF2C-AS1* | rs184944447 | 5 | 87817728 | 0 | 0.7 | 0.84 | 0.80 |
| *MEF2C-AS1* | rs188832904 | 5 | 87658662 | 0 | 0.5 | 0.82 | 0.79 |
| *MEF2C-AS1* | rs142862331 | 5 | 88015336 | 0 | 0.5 | 0.53 | 0.52 |
| *GABBR1-MOG* | rs75140056 | 6 | 29608184 | 0 | 1 | 1.03 | 1.00 |
| *GABBR1-MOG* | rs114178179 | 6 | 29606794 | 0 | 1 | 1.02 | 0.99 |
| *GABBR1-MOG* | rs115621016 | 6 | 29670478 | 0 | 0.6 | 1.03 | 1.00 |
| *GABBR1-MOG* | rs3131875 | 6 | 29666111 | 2 | 0.5 | 1.03 | 1.00 |
| *GABBR1-MOG* | rs71946391 | 6 | 29608870 | 0 | 0.5 | 0.91 | 0.90 |
| *GABBR1-MOG* | rs29269 | 6 | 29617747 | 2 | 0.5 | 1.01 | 1.00 |
| *ETV7* | rs9470264 | 6 | 36344980 | 0 | 1 | 1.06 | 0.96 |
| *ETV7* | rs3778028 | 6 | 36349666 | 0 | 1 | 1.09 | 0.99 |
| *ETV7* | rs1885206 | 6 | 36351095 | 0 | 0.9 | 1.09 | 0.99 |
| *ETV7* | rs113585688 | 6 | 36405876 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs76676074 | 6 | 36361063 | 0 | 0.8 | 1.12 | 0.99 |
| *ETV7* | rs113849189 | 6 | 36383209 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs16888788 | 6 | 36521517 | 0 | 0.8 | 1.13 | 0.99 |
| *ETV7* | rs4711451 | 6 | 36410308 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs2273883 | 6 | 36410801 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs111654791 | 6 | 36376021 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs80047437 | 6 | 36376369 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs4713977 | 6 | 36424664 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs4713976 | 6 | 36417410 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs112394821 | 6 | 36379855 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs79328581 | 6 | 36454464 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs79235676 | 6 | 36388019 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | chr6:36388436:I | 6 | 36388436 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs74555455 | 6 | 36389418 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs80331350 | 6 | 36391236 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs150014990 | 6 | 36392403 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs113005836 | 6 | 36519616 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs79927326 | 6 | 36434643 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs2273882 | 6 | 36437713 | 2 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs941974 | 6 | 36438385 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs75297867 | 6 | 36461335 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs4236051 | 6 | 36469821 | 2 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs2267930 | 6 | 36472788 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs75144098 | 6 | 36475146 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs113412960 | 6 | 36500799 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs111800900 | 6 | 36480859 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs3819759 | 6 | 36496277 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs111666164 | 6 | 36498423 | 0 | 0.8 | 1.12 | 1.00 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *ETV7* | rs112740031 | 6 | 36498559 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs75112107 | 6 | 36492839 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs113553570 | 6 | 36493878 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs2239541 | 6 | 36494053 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs201891379 | 6 | 36515576 | 0 | 0.8 | 1.11 | 0.99 |
| *ETV7* | rs12526415 | 6 | 36403874 | 2 | 0.8 | 1.11 | 1.00 |
| *ETV7* | rs200860324 | 6 | 36403871 | 0 | 0.8 | 1.10 | 0.99 |
| *ETV7* | rs116957783 | 6 | 36524594 | 0 | 0.8 | 1.11 | 0.99 |
| *ETV7* | rs112845031 | 6 | 36486403 | 0 | 0.8 | 1.12 | 1.00 |
| *ETV7* | rs78413663 | 6 | 36520101 | 0 | 0.7 | 1.11 | 1.00 |
| *ETV7* | rs76976387 | 6 | 36521363 | 0 | 0.7 | 1.11 | 1.00 |
| *ETV7* | rs202237535 | 6 | 36438384 | 0 | 0.7 | 1.01 | 0.90 |
| *ETV7* | rs4140594 | 6 | 36564240 | 0 | 0.7 | 1.09 | 0.99 |
| *ETV7* | rs3756907 | 6 | 36563129 | 0 | 0.7 | 1.09 | 0.99 |
| *ETV7* | rs149171383 | 6 | 36517137 | 0 | 0.7 | 0.96 | 0.87 |
| *ETV7* | rs7747978 | 6 | 36525321 | 0 | 0.7 | 1.06 | 0.97 |
| *ETV7* | rs7740345 | 6 | 36344329 | 0 | 0.6 | 0.92 | 0.84 |
| *ETV7* | rs4711454 | 6 | 36555422 | 0 | 0.6 | 1.08 | 1.00 |
| *ETV7* | rs3756906 | 6 | 36563812 | 0 | 0.6 | 1.08 | 1.00 |
| *ETV7* | rs76964964 | 6 | 36559290 | 0 | 0.6 | 1.08 | 1.00 |
| *ETV7* | chr6:36576246:D | 6 | 36576246 | 0 | 0.6 | 1.08 | 1.00 |
| *ETV7* | rs59108254 | 6 | 36576153 | 0 | 0.6 | 1.08 | 1.00 |
| *ETV7* | rs60633269 | 6 | 36576414 | 2 | 0.6 | 1.08 | 1.00 |
| *ETV7* | rs4713975 | 6 | 36395761 | 2 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs7382195 | 6 | 36621000 | 0 | 0.6 | 1.12 | 0.99 |
| *ETV7* | rs1998266 | 6 | 36358289 | 2 | 0.6 | 1.05 | 1.00 |
| *ETV7* | rs111624705 | 6 | 36413945 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs78373049 | 6 | 36421695 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs76019396 | 6 | 36425227 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs112542373 | 6 | 36416790 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs4711455 | 6 | 36632458 | 0 | 0.6 | 1.13 | 1.00 |
| *ETV7* | rs140429491 | 6 | 36422215 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs4713998 | 6 | 36632610 | 2 | 0.6 | 1.13 | 1.00 |
| *ETV7* | rs16888605 | 6 | 36433408 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs77660021 | 6 | 36431499 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs75776623 | 6 | 36477822 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs41272166 | 6 | 36455838 | 2 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs116413447 | 6 | 36479537 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs4713978 | 6 | 36466787 | 2 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs80032245 | 6 | 36492550 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs77693176 | 6 | 36548128 | 0 | 0.6 | 1.07 | 0.99 |
| *ETV7* | rs111747725 | 6 | 36509866 | 0 | 0.6 | 1.09 | 1.00 |
| *ETV7* | rs12525802 | 6 | 36445288 | 0 | 0.6 | 1.08 | 1.00 |
| *ETV7* | rs144398714 | 6 | 36346581 | 2 | 0.5 | 1.09 | 1.00 |
| *ETV7* | rs138977532 | 6 | 36349802 | 2 | 0.5 | 1.09 | 1.00 |
| *ETV7* | rs183032639 | 6 | 36517164 | 0 | 0.5 | 0.88 | 0.81 |
| *ETV7* | rs7758498 | 6 | 36344213 | 0 | 0.5 | 0.87 | 0.81 |
| *ETV7* | rs7776298 | 6 | 36548225 | 0 | 0.5 | 1.06 | 0.99 |
| *ETV7* | rs78502475 | 6 | 36315211 | 0 | 0.5 | 1.07 | 0.93 |
| *ETV7* | rs1744653 | 6 | 36397688 | 0 | 0.5 | 1.07 | 1.00 |
| *ETV7* | rs7741260 | 6 | 36399120 | 2 | 0.5 | 1.07 | 1.00 |
| *ETV7* | rs74693071 | 6 | 36298059 | 0 | 0.5 | 0.98 | 0.86 |
| *ACTN1* | rs117672662 | 14 | 69425467 | 0 | 1 | 1.01 | 0.98 |
| *ACTN1* | rs140754108 | 14 | 69407222 | 0 | 0.8 | 1.02 | 0.99 |
| *ACTN1* | rs12431622 | 14 | 69415032 | 0 | 0.8 | 1.01 | 0.98 |
| *ACTN1* | rs143493318 | 14 | 69369122 | 0 | 0.6 | 1.04 | 0.98 |
| *ACTN1* | rs3784133 | 14 | 69364953 | 2 | 0.5 | 1.06 | 1.00 |
| *ACTN1* | rs190190833 | 14 | 69350614 | 0 | 0.5 | 0.82 | 0.80 |
| *ACTN1* | rs76041256 | 14 | 69401213 | 0 | 0.5 | 1.01 | 1.00 |

[1]0=SNP was imputed,2= SNP was genotyped; [b]a statistical information metric from IMPUTE2

**References**


1. Gieger, C. *et al.* New gene functions in megakaryopoiesis and platelet formation. *Nature* **480**, 201-8 (2011).
2. Li, J. *et al.* GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children. *Hum Mol Genet* **22**, 1457-64 (2013).
3. Shameer, K. *et al.* A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. *Hum Genet* **133**, 95-109 (2014).
4. Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet* **41**, 1182-90 (2009).
5. Qayyum, R. *et al.* A meta-analysis and genome-wide association study of platelet count and mean platelet volume in african americans. *PLoS Genet* **8**, e1002491 (2012).
6. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* **42**, 210-5 (2010).
7. Martens, J.H. & Stunnenberg, H.G. BLUEPRINT: mapping human blood cell epigenomes. *Haematologica* **98**, 1487-9 (2013).
8. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
9. Tijssen, M.R. *et al.* Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell* **20**, 597-609 (2011).
10. Martens, J.H. *et al.* ERG and FLI1 binding sites demarcate targets for aberrant epigenetic regulation by AML1-ETO in acute myeloid leukemia. *Blood* **120**, 4038-48 (2012).
11. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (2013).
12. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
13. Ritchie, G.R., Dunham, I., Zeggini, E. & Flicek, P. Functional annotation of noncoding sequence variants. *Nat Methods* **11**, 294-6 (2014).
14. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-7 (2012).
15. Lee, D. *et al.* A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet* **47**, 955-61 (2015).