Published in final edited form as:
Hum Genet. 2014 January ; 133(1): . doi:10.1007/s00439-013-1355-7.

A Genome- and Phenome-Wide Association Study to Identify Genetic Variants Influencing Platelet Count and Volume and their Pleiotropic Effects

Khader Shameer[1], **Joshua C. Denny**[2], **Keyue Ding**[1], **Hayan Jouni**[1], **David R. Crosslin**[3], **Mariza de Andrade**[4], **Christopher G. Chute**[4], **Peggy Peissig**[5], **Jennifer A. Pacheco**[6], **Rongling Li**[7], **Lisa Bastarache**[2], **Abel N. Kho**[8], **Marylyn D Ritchie**[9], **Daniel R. Masys**[10], **Rex L. Chisholm**[6], **Eric B. Larson**[11], **Catherine A. McCarty**[12], **Dan M. Roden**[13], **Gail P. Jarvik**[14], and **Iftikhar J. Kullo**[1,§]

[1]Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55905, USA

[2]Departments of Medicine and Biomedical Informatics, Vanderbilt University, Nashville, TN 37232, USA

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

[4]Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN 55905, USA

[5]Biomedical Informatics Research Center, Marshfield Clinic, Marshfield, WI, 54449, USA

[6]Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA

[7]Office of Population Genomics, National Human Genome Research Institute, 5635 Fishers Lane, Suite 3058, MSC 9307, Bethesda, MD, 20892, USA

[8]Department of Medicine, Northwestern University, Chicago, IL 60611, USA

[9]Center for Systems Genomics, Pennsylvania State University, Eberly College of Science, The Huck Institutes of the Life Sciences, 512 Wartik Laboratory, University Park, PA 16802 USA

[10]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Room 416 Eskind Medical Library, Nashville, TN, 37232, USA

[11]Group Health Research Institute, 1730 Minor Avenue, Suite 1600, Seattle, WA, 98101, USA

[12]Essentia Institute of Rural Health, Duluth, MN, 55805 USA

[13]Department of Pharmacology, Vanderbilt University School of Medicine, 1285 Medical Research Building IV, Nashville, TN, 37232, USA

[14]Department of Genome Sciences, University of Washington, 3720 15th Ave NE, Seattle WA 98195, USA

Abstract

Platelets are enucleated cell fragments derived from megakaryocytes that play key roles in hemostasis and in the pathogenesis of atherothrombosis and cancer. Platelet traits are highly heritable and identification of genetic variants associated with platelet traits and assessing their pleiotropic effects may help to understand the role of underlying biological pathways. We conducted an electronic medical record (EMR)-based study to identify common variants that

[§]**Corresponding author:** Iftikhar J. Kullo, M.D., 200 First Street SW, Rochester, MN 55905, Kullo.iftikhar@mayo.edu, Phone: 507-284-9049.

Competing financial interests The authors declare no competing financial interests.

influence inter-individual variation in the number of circulating platelets (PLT) and mean platelet volume (MPV), by performing a genome-wide association study (GWAS). We characterized association of variants influencing MPV and PLT using functional, pathway and disease enrichment analysis assess pleiotropic effects of such variants by performing a phenome-wide association study (PheWAS) with a wide range of EMR-derived phenotypes. A total of 13,582 participants in the electronic MEdical Records and GEnomic (eMERGE) network had data for PLT and 6,291 participants had data for MPV. We identified 5 chromosomal regions associated with PLT and 8 associated with MPV at genome-wide significance (P<5E-8). In addition, we replicated 20 SNPs (out of 56 SNPs (α: 0.05/56=9E-4)) influencing PLT and 22 SNPs (out of 29 SNPs (α: 0.05/29=2E-3)) influencing MPV in a meta-analysis of GWAS of PLT and MPV. While our GWAS did not reveal any novel associations, our functional analyses revealed that genes in these regions influence thrombopoiesis and encode kinases, membrane proteins, proteins involved in cellular trafficking, transcription factors, proteasome complex subunits, proteins of signal transduction pathways, proteins involved in megakaryocyte development and platelet production and hemostasis. PheWAS using a single-SNP Bonferroni correction for 1368 diagnoses (0.05/1368=3.6E-5) revealed that several variants in these genes have pleiotropic associations with myocardial infarction, autoimmune and hematologic disorders. We conclude that multiple genetic loci influence interindividual variation in platelet traits and also have significant pleiotropic effects; the related genes are in multiple functional pathways including those relevant to thrombopoiesis.

Introduction

Platelets or thrombocytes are enucleated cell fragments derived from precursor megakaryocytes during megakaryopoiesis. Platelets play a key role in hemostasis and in the pathophysiology of atherothrombosis and cancer (1–8). In addition, platelets are involved in several biological processes including immune and inflammatory response, vascular integrity, wound healing, and tumor metastasis (9–13).

The number of circulating platelets (PLT) in an individual is regulated by a balance between thrombopoiesis and platelet consumption and senescence (14). The mean platelet volume (MPV), a measure of platelet size, is determined at the time of platelet production from megakaryocytes and is inversely related to PLT. The precise molecular mechanisms regulating PLT and MPV are not well characterized. Platelet traits are heritable and in one study (15), 84% of the variance in PLT was attributed due to genetic factors. A high heritability of 75% for MPV has also been reported (9,16,17).

To identify genetic variants associated with PLT and MPV we performed a genome-wide association study (GWAS) in 13,582 participants of European ancestry in the electronic MEdical Records and GEnomic (eMERGE) network (http://www.gwas.net). As platelets and platelet-derived growth factors are important mediators in inflammation, atherosclerosis, and neoplasia (13,18–21), we investigated whether genetic variants associated with PLT/MPV have pleiotropic effects on other disease phenotypes, using a phenome-wide association study (PheWAS) (22,23). (Figure S1). The PheWAS approach is uniquely possible in the eMERGE network given that the genotypes are linked to wide range of phenotypes represented in the electronic medical record (EMR). Further, we inferred the functional categories, biological pathways and diseases mediated by genes encoded in the loci associated with PLT and MPV using biological enrichment analyses.

Methods

Phenotyping

The complete blood count (CBC), typically measured by automated hematology analyzers is a commonly performed test in the clinical setting. PLT values were available in the EMR for up to 13, 582 patients with extant GWAS data from the five sites of the eMERGE I network (24) as part of the CBC; MPV values were available for 46.3% (n=6,291) of the patients in the cohort. Samples with PLT values ≤ 100 or ≥ 600 (×109/L) were excluded. Of the remaining PLT and MPV lab values, the median value was extracted for each study participant and included in further analyses. Additionally, we excluded patients with hematological disorders including cancer, iron deficiency and anemia using an EMR-based algorithm (available at the compendium of EMR derived clinical phenotypes (PheKB; URL: http://phekb.org/)); The algorithm for PLT and MPV traits can be accessed at the URL: http://www.phekb.org/phenotype/platelet-count-plt-and-mean-platelet-volume-mpv.

EMR-linked GWAS of PLT & MPV

We performed a GWAS for PLT in 13,582 patients of European ancestry identified from five sites in the eMERGE network (Table 1); and for MPV in 6, 291 patients from Mayo Clinic, Marshfield Clinic, and Northwestern University. The study design (Figure S1) and the distributions of PLT and MPV traits are depicted in the supplemental material (Figure S2). The clinical laboratories of Group Health and Vanderbilt did not report MPV as part of a CBC. As previously described (25,26), genotyping was performed in patients of European ancestry at the Center for Genotyping and Analysis at the Broad Institute (MAYO, VUMC, and NU) and the Center for Inherited Disease Research at the Johns Hopkins University (MC and GHC) using the Illumina Human660W-Quadv1_A genotyping platform, consisting of 561,490 SNPs and 95,876 intensity-only probes. Data were cleaned using the quality control (QC) pipeline developed by the eMERGE Genomics Working Group (26). The process includes evaluation of sample and marker call rate, gender mismatch and anomalies, duplicate and HapMap concordance, batch effects, Hardy-Weinberg equilibrium, sample relatedness, and population stratification. A total of 476,395 SNPs were available for analysis after applying the following QC criteria: SNP call rate >98%, sample call rate >98%, minor allele frequency >0.05, Hardy-Weinberg equilibrium *P*-value >0.001, and 99.99% concordance rate in duplicates. The first two dimensions were derived from multidimensional decomposition analysis (using ‘cmdscale’ function in *R*) of the 1-IBS (identity-by-state) matrix based on all eMERGE samples (*n*=17,358). Samples >6 standard error from the mean of self-reported white ethnicity on dimensions 1 and 2 were excluded. After QC steps, 13,582 patients (i.e., patients with genetically-defined European ancestry) with phenotype and genotype data were available for association analyses.

Statistical analyses

When multiple measurements of platelet traits were available for an individual patient, we chose the median value and the corresponding age for the genetic analyses. We used an efficient mixed-model association expedited (EMMAX) algorithm (27) to correct for sample relatedness and cryptic population substructure. The IBS matrix was calculated for each pair of individuals using the genome-wide genotype data. The generalized least squares F-test was used to estimate the regression coefficient () and perform association analyses, which were implemented in the R ‘emmax’ package, with adjustment for age, sex, and site (28). The statistical power of our study was 80% to detect a quantitative trait locus that explained 0.29% of variance in PLT and 0.63% in MPV, given a sample size of 13,582 and 6,291 respectively and a significance level of 5E-8. We imputed all non-genotyped SNPs in 20 autosomal chromosomes based on HapMap II CEU database (release 21). Imputation-based association for non-genotyped SNPs was performed using the same IBS matrix for the

genotyped SNPs by EMMAX. Locuszoom (29) was used for the visualization of genomic loci associated with platelet traits.

Imputation analysis

A recent genome-wide meta-analyses of PLT (*n*=48,000) and MPV (*n*=23,000) in subjects of European ancestry found 85 SNPs in 69 regions on 20 autosomal chromosomes that were associated with PLT and MPV (30). Of the 56 SNPs associated with PLT, 23 were genotyped on the Illumina 660W platform used in the eMERGE network. We were able to impute the remaining 33 based on HapMap II CEU using the Markov Chain based haplotyper (MACH) program (31).

PheWAS of variants influencing PLT & MPV

We performed PheWAS on 81 SNPs achieving genome-wide significant associations with PLT, MPV, or both, either in this study or in a prior GWAS of PLT and MPV (30). For this analysis, we combined the entire eMERGE cohort of European American individuals (n=13, 688) identified across the five eMERGE sites. To define diseases, we queried all International Classification of Disease (ICD), 9th edition, codes from the respective EMRs of the five eMERGE sites as previously described (32). Briefly, for each disease, the PheWAS algorithm constructed case populations by grouping ICD codes into 1368 related case groups (e.g., “type 1 diabetes”, “hypertension”). To be a case, an individual must have at least 2 ICD codes in that case group. A control population for each phenotype is constructed by selecting all patients that do not have the case disease or closely related diseases (e.g., a patient with a bundle branch block cannot serve as a control for complete heart block). Based on a single-SNP Bonferroni correction, a *P-value* of 0.05/1368=3.6E-5 was considered to be significant. The PheWAS methodology (22) has previously been validated through rediscovery of known associations (33). Analysis of each phenotype then proceeds using a pairwise analysis of all case and control groups for each tested SNP. As with other PheWAS, we did not analyze phenotypes that occurred in less than 25 patients (a prevalence of 0.18% in the dataset). The PheWAS was performed with PLINK using logistic regression analyses that adjusted for age and gender and assumed an additive genetic model (34).

Functional enrichment analysis of genes harboring variants regulating PLT/MPV

GWAS analysis implicated multiple genetic variants as modulators of platelet traits. To understand the functional similarity and diversity of these genes, we used a bioinformatics analysis pipeline (See Supplementary material; Figure S2) that combined 3 different tools: Protein ANalysis Through Evolutionary Relationship (PANTHER) (35) was used to perform Gene Ontology term and protein class enrichment analysis. These analyses were used to understand the functional similarity of genes associated with PLT and MPV. GO term enrichment analyses were used to find significant (binomial statistics; $P<0.05$) terms under the categories of biological processes, molecular functions and cellular compartments. Protein class enrichment analyses were performed (binomial statistics; $P<0.05$) to find evolutionarily conserved globular domains encoded in genes influencing platelet traits. Reactome pathway annotations (36) were used to identify major molecular events mediated by the genes. Significance was defined using a hypergeometric test ($P<0.05$). Functional annotation based on Gene Ontology or molecular event annotations does not capture known gene-disease associations. To identify the disease-associated genes influencing PLT and MPV, we performed an enrichment analysis using Disease Ontology. Disease Ontology enrichment analysis was performed using Functional Disease Ontology server using a simplified version of Disease Ontology (DO-Lite) (37) to identify enrichment of diseases reported to be associated with genes influencing PLT and MPV (hypergeometric test;

P<0.05). Expression levels of genes harboring the variants that regulate PLT and MPV in megakaryocyte (*n*=4) were examined using HaemAtlas (38).

Results

There were 13424 patients of European ancestry with PLT data; 6, 291 of these patients also had MPV data from five eMERGE sites after exclusion of patients (n=158) with with abnormal PLT or MPV values. Sample characteristics are shown in Table 1; distributions of platelet traits across the eMERGE sites are provided in Figure S2. The two traits were negatively correlated (gender-adjusted r=−0.36) with each other. We identified 5 chromosomal regions associated with PLT (Figure 1a), 8 chromosome regions associated with MPV (Figure 1b) (*P* <5E-8) (Table 2) and 1 region (3p14.3) associated with both traits.

Regions associated with PLT

Chromosome 3p14.3—An intronic SNP (rs1354034) in the Rho guanine nucleotide exchange factor 3 gene (*ARHGEF3*) was associated with PLT (=7.97, *P*=6E-24). Rho-like GTPases are involved in a variety of cellular processes, activated by binding GTP, and inactivated by conversion of GTP to GDP by their intrinsic GTPase activity.

Chromosome 12q24.12—A missense SNP (rs3184504) in *SH2B3* was associated with PLT (=−5.33, *P*=5E-12). This SNP has been associated with multiple phenotypes, including type 1 diabetes, (39,40) blood pressure, (41) eosinophil count and myocardial infarction (42). Another SNP (rs739496, r2=0.23 with rs3184504) at this locus was associated with PLT in a Japanese cohort (43). Soranzo et al. (44) found an adjacent SNP (rs11065987) in *ATXN2* to be associated with PLT. We also found a SNP rs653178 in *ATXN2* to be associated with PLT, however this SNP is in LD with the missense SNP rs3184504 (r^2=1) in *SH2B3*. The pleiotropic associations of rs3184504 reported in prior studies along with findings from PLT/MPV GWAS-PheWAS analyses are summarized in Table S8.

Chromosome 11p15.5—Several SNPs near the telomere region of chromosome 11p were associated with PLT (the most significant SNP, rs11602954, =−6.46, *P*=5E-12). Six genes reside in this region: *SCGB1C1*, *ODF3*, *RIC8A*, *BET1L*, *SIRT3*, and *PSMD13*. The SNP rs11602954 may regulate expression of the neighboring genes *BET1L* and *SIRT3* as well as *PSMD13* to a lesser degree (44).

Chromosome 6q23.3—An intergenic SNP (rs4895441) between *HBS1L* and *MYB* genes was associated with PLT (=−5.24, *P*=9 10E-10). Another intergenic SNP (rs9399137, *P*=1 10^{-9}) within this region was also associated with PLT. (45) We have previously found this locus to be also associated with red blood cell (RBC) traits: RBC count, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH) and mean corpuscular hemoglobin (MCHC) (25).

Chromosome 9p24.1—A SNP in the 5’ flanking region of *RCL1* was associated with PLT (=−4.94, *P*=1E-9). The locus was noted to be to be associated with PLT (rs385893, *P*=9E-17) in a previous study (44). Regional plots of five chromosomal regions influencing PLT are provided in Figure 2.

Regions associated with MPV

Chromosome 3p14.3—An intronic SNP (rs1354034) in *ARHGEF3* was associated with MPV (=−0.19, *P*=9E-34). SNPs in *ARHGEF3* have been reported to be associated with MPV in previous studies (44,46).

Chromosome 7q22.3—Several SNPs on chromosome 7q22.3 were associated with MPV (most significant SNP rs342240, =−0.15, *P*=5E-22). These SNPs are located in the intergenic region of *FLJ36031* and *PIK3CG*. Soranzo et al (44) found this locus to be associated with MPV (*P*=1E-24). RNA expression analysis of this region showed a significant association with *PIK3CG* expression and platelet levels (47). This locus contains multiple transcription factor binding sites that regulate the expression level of *PIK3CG* and rs342240 disrupts an *in-silico* predicted transcription factor binding site in the intergenic region (48).

Chromosome 1q32.1—Multiple SNPs in LD on chromosome 1q32.1 were associated with MPV (most significant SNP rs9660992, =0.11, *P*=3E-13). This region contains several genes including *CNTN2*, *TMEM81*, *DSTYK*, *TMCC2*, and *NUAK2*. The SNP rs9660992 is located in the 3' untranslated region of *TMCC2*. Another intronic SNP in *TMCC2* (rs1668873, r^2=0.62 with rs9660992) has been reported to be associated with MPV (44).

Chromosome 1q24.3—Intronic SNPs in *DNM3* were associated with MPV (the most significant SNP rs2180748, =0.09, *P*=2E-8). An intronic SNP rs10914144 in *DNM3* (*P*=2E-14), which is in moderate LD (r^2=0.31) with rs2180748, was associated with MPV in a previous study (44).

Chromosome 10q21.2—Multiple SNPs in LD, spanning a region of ~400 kb on chromosome 10q21.2 (lead SNP: rs4379723 P=3E-16) were associated with MPV. Five genes reside in this region, including *NRBF2*, *JMJD1C*, *MIR1296*, *LOC84989*, and *REEP3*. This region has been previously reported to be associated with MPV (44).

Chromosome 17q11.2—Three intronic SNPs in *TAOK1* and one SNP in the 3' regulatory region of *TAOK1* on chromosome 17q11.2 were associated with MPV (the most significant SNP rs9900280, =0.1, *P*=1E-10). This locus has been reported to be associated with MPV in previous studies (44,46).

Chromosome 12q13.13—Two SNPs in LD on chromosome 12q13.13 were associated with MPV. The most significant SNP (rs10506328, =−0.09, *P*=2E-9) is in an intron of *NFE2*. The same SNP was reported to be associated with MPV (*P*=5E-7) in a previous study (44).

Chromosome 12q24.1—Multiple SNPs on chromosome 12q24.1 were associated with MPV. The most significant SNP rs7961894 (=−4.94, *P*=1E-9) is located in an intron of *WDR66*. Variants in *WDR66* have been reported to be associated with MPV (46,47). The LD between rs7961894 and other significant SNPs was weak (r^2 <0.3) and it is possible that multiple independent genes contribute to variation in MPV including *TMEM120B*, *SETD1B*, *PSMD9*, and *BCL7A*. Regional plots of 8 chromosomal regions influencing PLT are provided in Figure 3.

Replication of previously reported genetic regions associated with PLT/MPV

Using an alpha of 9E-04 to correct for multiple testing, nine of the 23 genotyped and 11 of the 33 imputed SNPs were replicated (Table S1). Of 29 SNPs associated with MPV, five were genotyped in the present study and 24 were imputed. A total of 22 SNP associations with MPV were replicated (Table S2). Genes in these regions influence thrombopoiesis and encode kinases (*AK3*, *DSTYK*, *PIK3CG*, *TAOK1*), membrane proteins (*BET1L*, *COPZ1*, *REEP3*, *TMCC2*, *TMEM81*, *TMEM120B*), proteins involved in cellular trafficking (*DNM3*, *NRBF2*, *REEP3*), transcription factors (*JMJD1C*, *MYB*, *NFE2*, *NRBF2*, *RBBP5)*, proteasome complex subunits (*PSMD13*), proteins of signal transduction pathways

(*ARHGEF3, SH2B3*), proteins involved in megakaryocyte development and platelet production (*AK3, CABLES1, EHD3, JMJD1C, MYB, NFE2, SH2B3*) and hemostasis (*SH2B3, JMJD1C, MYB, EHD3, PIK3CG, F2R, SIRPA, PTPN11, CABLES1, AK3, THPO, NFE2*).

A recent meta-analysis of GWAS in African Americans (49) revealed 11 regions to be associated with PLT and MPV. We replicated five loci associated with PLT (6q23, 7q22, 10q21, 12q24) and 2 loci (7q22 and 17q11) associated with MPV, in the European ancestry patients in our cohort.

PheWAS of variants associated with PLT and MPV

We analyzed a total of 81 SNPs using PheWAS: 56 for PLT, 29 for MPV (4 SNPs were associated with both phenotypes). Phenotypic associations are summarized in Table 3. Figure 4 shows the PheWAS plots for SNPs associated with PLT and MPV. Multiple autoimmune and hematological conditions were associated with these SNPs. The strongest associations were of ankylosing spondylitis (P=3.3E-7), inflammatory spondylopathies (P=5.7E-8), and uveitis (P=4.6E-7) with rs381299 in *HLA-B*. Associations between *HLA-B* and ankylosing spondylitis have been reported previously (50). Additional associations included that of psoriasis and psoriatic arthropathy with rs6065 in *GP1BA*, dermatomyositis/polymyositis with rs7149242 at 14q32.2 near C14orf70/*DLK1*, and hypothyroidism (which is typically an autoimmune disease) with missense variant rs3184504 in *SH2B3*, all at P<2E-4. A borderline significant association was noted for myocardial infarction and this missense variant in *SH2B3* (P=1.7E-04). Several GWAS have noted the association between this missense variant (rs3184504) with several complex traits and disease phenotypes including hypothyroidism (51) and myocardial infarction (42) (Table S8). The mutation results in substitution of a nonpolar and neutral residue with a polar and positive residue on the pleckstrin homology (PH) domain of *SH2B3* leading to an additional positively charged residue to the region that may influence the binding of *SH2B3* with other protein partners involved in blood coagulation, hemostasis or platelet activation pathways. Additional associations near statistical significance (all P<1.2E-04) noted in PheWAS were those of several hematological conditions: B12-deficiency anemia with rs9399137 at 6q23.3 near *HBS1L– MYB*, thrombocytopenia with rs2950390 at 12q13.3 near *PTGES3–BAZ2A*, leukemia with rs6995402 in plectin gene (*PLEC*), and aplastic anemia with rs7961894 in WD repeat-containing protein 66 (*WDR66*).

Biological enrichment analysis of genes influencing PLT & MPV

We investigated functional associations of genes influencing PLT and MPV. Detailed descriptions of the methods (Figure S3) are provided in the Supplementary Material. Briefly, biological enrichment analysis was performed using a merged list of genes from the significant loci identified from GWAS and imputation analysis. We identified significantly enriched GO annotations using enrichment analysis. GO term analyses (Table S3) revealed seven terms under ‘biological processes’ category that were associated with PLT and MPV; out of which five terms were significant for both traits. Six terms under ‘molecular function’ category were associated with PLT and five terms were associated with MPV with one term “lyase activity” associated with genes influencing both traits. Protein classes (Table S4) associated with the genes were ascertained using protein class enrichment analysis using protein relationships compiled in PANTHER Ontology. Protein classes “membrane traffic protein” and “vesicle coat protein” were associated with genes influencing both PLT and MPV. Molecular events (individual cellular events derived from biological processes delineated as molecular interactions facilitated by enzymes or other molecular catalysts) associated with genes were derived using Reactome annotations (Table S5). Two molecular events “Factors involved in megakaryocyte development and platelet production” and

"Hemostasis" were associated with both traits. Disease Ontology terms (Figure S4 and Table S6) associated with genes influencing PLT and MPV were derived. Five Disease Ontology terms including "myeloproliferative disease", "leukemia" and "autoimmune disease" were identified as significant in disease term enrichment analysis. GO term enrichment analysis did not reveal any significant cellular compartment terms. Functional in vitro studies on GWAS using *in-vitro* studies have shown that specific genetic variants have critical role in cell-types (52–54). To understand cell-type specific expression profile of platelet trait-associated genes, we leveraged publicly available expression data. Expression levels of genes influencing PLT and MPV in megakaryocytes (precursor of platelets in hematopoietic cell lineage) were obtained from HaemAtlas database; we noted several genes had high level of expression in the megakaryocytes (See Table S7 and Figure S3). We also performed a comparative analysis of genes implicated in platelet function, PLT and MPV and noted SNPs in *JMJD1C* to be associated with all three traits (See Supplementary Material; Figure S5)

Discussion

In the present study, we performed an EMR-based GWAS of PLT and MPV in 13,582 patients of European ancestry from five sites of the eMERGE network and examined the pleiotropic effects of these variants using a PheWAS approach. We identified five chromosome regions associated with PLT and eight associated with MPV respectively. We replicated 20 SNPs influencing PLT (out of 56 SNPs (α: 0.05/56=9E-04)) and 22 SNPs (out of 29 SNPs (α: 0.05/29=2E-03)) influencing MPV from among 69 SNPs discovered in a meta-analysis of PLT and MPV (30).

PheWAS analysis using a single-SNP Bonferroni correction (0.05/1368=3.6E-5) revealed that several variants influencing PLT and MPV have pleiotropic effects and are associated with diverse phenotypes including myocardial infarction, autoimmune and hematologic diseases, highlighting pleiotropic effects of these variants. Additionally, several variants were associated at near-significance with hematologic conditions that may, in part, explain their association with PLT and/or MPV. For instance, PheWAS replicated the known association between rs3819299 (an intronic variant of *HLA-B* associated with PLT) and ankylosing spondylitis. Indeed, MPV has been used as a marker of disease activity in ankylosing spondylitis and rheumatoid arthritis (55). Our study illustrates the potential use of EMR-based genomic studies to identify novel genetic loci for medically important quantitative traits, as well as to replicate prior associations and identification of potential genotype-phenotype association using the PheWAS approach.

A missense variant (rs3184504) in *SH2B3*, associated with PLT, was also associated with hypothyroidism and myocardial infarction in our PheWAS. This variant has been previously noted to be associated with cardiovascular diseases including myocardial infarction and coronary disease (56,57), Type 1 diabetes (58), autoimmune disorders (multiple sclerosis, systemic lupus erythematous, Graves disease, juvenile rheumatoid arthritis, Addison disease and celiac disease) (59,60), asthma (42), hypothyroidism (51) and hypertension (61). (See Table S8 for a summary of the diverse phenotypes associated with rs3184504). *SH2B3* encodes a multi-domain protein with two highly conserved functional domains (PH domain and SH2 (Src homology 2) domain) and is involved in blood coagulation and erythropoietin (EPO) signaling pathway (62). The variant rs3184504 influencing PLT is a missense mutation on *SH2B3* gene resulting in a residue change from tryptophan (large size and aromatic side chain) to arginine (large size and basic side chain) substitution. This variant is present in a conserved PH domain involved in lipid binding (63) and was found to be associated with hypothyroidism and myocardial infarction in our PheWAS.

PheWAS analysis replicated the known association between rs3819299 (an intronic variant of *HLA-B* associated with PLT) and ankylosing spondylitis. Indeed, MPV has been used as a marker of disease activity in ankylosing spondylitis and rheumatoid arthritis (55). Several of the PheWAS findings are novel, representing rare conditions not yet studied in GWAS (e.g., dermatomyositis and chronic glomerulonephritis). The association of aplastic anemia with rs7961894 in *WDR66* suggests that the changes in PLT and MPV associated with these SNPs represent a continuum between variation of normal and conditions resulting in disease manifestation. Finally, another class of associations present may represent regulation of platelet count or function: menstrual irregularities, gastrointestinal complications, and complications of trauma. We also noted that several previously reported genotype-disease associations identified using GWAS were not captured in our PheWAS analysis. Additional studies are needed to validate these associations including studies in model organisms, epigenomic assays and targeted ex-vivo or in-vitro assays (64).

Statistical enrichment analyses were performed to understand functional repertoire, biological pathways and diseases mediated by genes regulating PLT and MPV. Functional enrichment analysis revealed that the genes influencing PLT and MPV traits included kinases, membrane proteins, transporter proteins including proteins involved in cellular trafficking, proteasome complex subunit, transcription factors as well as genes implicated in signal transduction pathways relevant to megakaryocyte development and platelet production. The platelet life cycle involves a series of complex biological processes and molecular functions. Variants influencing PLT and MPV are distributed in several genes that influence multiple cellular mechanisms such as trafficking, cell signaling pathways, membrane dynamics and generation of precursor cells of platelets (megakaryocytes). Previous investigations have reveled cellular trafficking mechanisms to be associated with platelet lifecycle including endocytosis /pinocytosis by megakaryocytes, phagocytosis of platelets by macrophages in the spleen, and secretion of platelet granules by exocytosis to mediate hemostasis (11,12,30,65–75). Signaling pathways have also shown to be pivotal for normal platelet function; platelet activation, aggregation, and adhesion are mediated by signal transduction events facilitated by signaling molecules and platelet receptors (76–78). Signal transduction pathways via phosphorylation events also influence platelet lifecycle (79,80). The ubiquitin-and proteasome-dependent proteolytic pathway degrades the majority of cellular proteins and previous studies indicate a role of proteolysis in megakaryocytes and platelet function (81,82).

Several genes involved in aforementioned molecular pathways were found to be harboring variants influencing PLT and MPV (Table S3, S4 and S5). For example, three protein kinases: adenylate kinase 3 (*AK3*), phosphoinositide-3-kinase catalytic subunit gamma isoform (*PIK3CG*) and TAO kinase 1 gene (*TAOK1*) encoding a microtubule affinity-regulating kinase that regulates mitotic progression and autophagy (83), were associated with platelet traits (Table S4). Protein kinases play a significant role in signal transduction via phosphorylation (79,80). Signal transduction and kinase related terms were significantly enriched among the genes influencing the two platelet traits (Tables S3 and S4; all $P<0.05$). Transcription factors such as nuclear factor (erythroid-derived 2) (*NFE2*) play an important role in platelet biogenesis. *NFE2* overexpression is associated with enhanced megakaryocyte differentiation, proplatelet development, and platelet release while loss of NFE2 function results in lethal embryonic thrombopoietic defects and severe thrombocytopenia (44,47). We also replicated multiple intergenic SNPs associated with PLT (*HBS1L-MYB* (84,85), *RPH3A– PTPN11* (30)) and MPV (*FLJ36031-PIK3CG* (86), *CD9–VWF* (30)) suggesting transcriptional regulation of platelet phenotypes. Further we observed that transcription factor related terms (translation elongation factor activity (P=8.11E-03), transcription cofactor activity (P=8.11E-03)) enriched among the genes that harbor variants influencing

platelet traits. Genes and related key functional pathways derived from functional annotation analyses are summarized in the context of platelet biogenesis in Figure 5.

The catalog of GWAS-identified genetic variations and their phenomic associations is expanding (33,87,88). Our PheWAS results highlighted novel phenome-associations of platelet related genetic variants such as dermatomyositis and polymyositis and chronic glomerulonephritis. While the PheWAS approach (22) has been validated through rediscovery of known associations, akin to any genetic or phenomic studies, replication and functional validation of the associations will be needed, particularly since our threshold for statistical significance in these analyses may be considered liberal.

Summary

Our study highlights the way in which high-throughput EMR-based genomic and phenomic studies can be integrated to identify genetic variants influencing clinically relevant traits and their pleiotropic effects. We identified five chromosomal regions associated with PLT, eight regions associated with MPV and one region associated with both traits. We interpreted GWAS results with functional repertoire analysis and integrated with PheWAS analysis to validate the disease phenotypes associated with variants influencing platelet phenotypes. The functional annotation approach that we employed using GO annotations, protein class annotations and molecular event annotations, highlights the diverse biological functions played by various gene families associated with platelet traits. Disease ontology analysis revealed disease context of six genes (*BAK1*, *CBL*, *HLA-B*, *PTPN11*, *MYB* and *TPM1*) influencing PLT and MPV and further investigation of these genes may help understand their contributions to platelet biology, thrombopoiesis and diverse disease pathways influenced by platelets. PheWAS analysis suggests that several SNPs influencing platelet traits are associated with diverse phenotypes including myocardial infarction, autoimmune diseases and hematologic disorders. Using an integrated bioinformatics pipeline, we demonstrate that the genes harboring variants regulating platelet traits include gene families that influence several aspects of thrombopoiesis. Our results provide a framework to understand the role of biological pathways in regulating platelet traits and the pleiotropic effects of genes influencing these traits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The eMERGE network was initiated and funded by the National Human Genome Research Institute (NHGR1), with additional funding from National Institute of General Medical Sciences (NIGMS) through the following grants: U01-HG-04599 (Mayo Clinic); U01-HG-004610 and UO1-AG-06781 (Group Health Cooperative); U01-HG-004608 (Marshfield Clinic); U01HG004609 (Northwestern University); and U01-HG-04603 (Vanderbilt University, also serving as the Administrative Coordinating Center). We also acknowledge the genotyping centers U01-HG-004424 (Broad Institute) and U01-HG-004438 (Johns Hopkins University, Center for Inherited Disease Research). Additional genotyping support was provided by a Washington State Life Sciences Discovery Fund award to the Northwest Institute of Genetic Medicine (G.P.J). Additional support for PheWAS was provided through R01-LM-010685.

References

1. Weber C. Platelets and chemokines in atherosclerosis: partners in crime. Circ Res. 2005; 96:612–616. [PubMed: 15802619]
2. Willoughby S, Holmes A, Loscalzo J. Platelets and cardiovascular disease. Eur J Cardiovasc Nurs. 2002; 1:273–288. [PubMed: 14622657]

3. Cameron HA, Phillips R, Ibbotson RM, Carson PH. Platelet size in myocardial infarction. Br Med J (Clin Res Ed). 1983; 287:449–451.
4. del Zoppo GJ. The role of platelets in ischemic stroke. Neurology. 1998; 51:S9–S14. [PubMed: 9744824]
5. McBane RD 2nd, Karnicki K, Miller RS, Owen WG. The impact of peripheral arterial disease on circulating platelets. Thromb Res. 2004; 113:137–145. [PubMed: 15115669]
6. Honn KV, Tang DG, Crissman JD. Platelets and cancer metastasis: a causal relationship? Cancer Metastasis Rev. 1992; 11:325–351. [PubMed: 1423821]
7. Pain A, Ferguson DJ, Kai O, Urban BC, Lowe B, Marsh K, Roberts DJ. Platelet-mediated clumping of Plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. Proc Natl Acad Sci U S A. 2001; 98:1805–1810. [PubMed: 11172032]
8. Kottke-Marchant K. Importance of platelets and platelet response in acute coronary syndromes. Cleve Clin J Med. 2009; 76(Suppl 1):S2–S7. [PubMed: 19332590]
9. Kunicki TJ, Nugent DJ. The genetics of normal platelet reactivity. Blood. 2010; 116:2627–2634. [PubMed: 20610812]
10. Kunicki TJ, Williams SA, Salomon DR, Harrison P, Crisler P, Nakagawa P, Mondala TS, Head SR, Nugent DJ. Genetics of platelet reactivity in normal, healthy individuals. J Thromb Haemost. 2009; 7:2116–2122. [PubMed: 19740098]
11. Smyth SS, McEver RP, Weyrich AS, Morrell CN, Hoffman MR, Arepally GM, French PA, Dauerman HL, Becker RC. Platelet functions beyond hemostasis. J Thromb Haemost. 2009; 7:1759–1766. [PubMed: 19691483]
12. Nachman RL, Rafii S. Platelets, petechiae, and preservation of the vascular wall. N Engl J Med. 2008; 359:1261–1270. [PubMed: 18799560]
13. Gay LJ, Felding-Habermann B. Contribution of platelets to tumour metastasis. Nat Rev Cancer. 2011; 11:123–134. [PubMed: 21258396]
14. Daly ME. Determinants of platelet count in humans. Haematologica. 2011; 96:10–13. [PubMed: 21193429]
15. Johnson AD. The genetics of common variation affecting platelet development, function and pharmaceutical targeting. J Thromb Haemost. 2011; 9(Suppl 1):246–257. [PubMed: 21781261]
16. Mahaney MC, Brugnara C, Lease LR, Platt OS. Genetic influences on peripheral blood cell counts: a study in baboons. Blood. 2005; 106:1210–1214. [PubMed: 15870178]
17. O'Donnell CJ, Larson MG, Feng D, Sutherland PA, Lindpaintner K, Myers RH, D'Agostino RA, Levy D, Tofler GH. Genetic and environmental contributions to platelet aggregation: the Framingham heart study. Circulation. 2001; 103:3051–3056. [PubMed: 11425767]
18. Gawaz M, Langer H, May AE. Platelets in inflammation and atherogenesis. J Clin Invest. 2005; 115:3378–3384. [PubMed: 16322783]
19. Huo Y, Ley KF. Role of platelets in the development of atherosclerosis. Trends Cardiovasc Med. 2004; 14:18–22. [PubMed: 14720470]
20. Gay LJ, Felding-Habermann B. Platelets alter tumor cell attributes to propel metastasis: programming in transit. Cancer Cell. 2011; 20:553–554. [PubMed: 22094248]
21. Labelle M, Begum S, Hynes RO. Direct Signaling between Platelets and Cancer Cells Induces an Epithelial-Mesenchymal-Like Transition and Promotes Metastasis. Cancer Cell. 2011; 20:576–590. [PubMed: 22094253]
22. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010; 26:1205–1210. [PubMed: 20335276]
23. Pendergrass SA, Brown-Gentry K, Dudek SM, Torstenson ES, Ambite JL, Avery CL, Buyske S, Cai C, Fesinmeyer MD, Haiman C, et al. The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. Genet Epidemiol. 2011; 35:410–422. [PubMed: 21594894]
24. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, Li R, Masys DR, Ritchie MD, Roden DM, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics. 2011; 4:13. [PubMed: 21269473]

25. Kullo IJ, Ding K, Jouni H, Smith CY, Chute CG. A genome-wide association study of red blood cell traits using the electronic medical record. PLoS One. 2010; 5
26. Turner, S.; Armstrong, LL.; Bradford, Y.; Carlson, CS.; Crawford, DC.; Crenshaw, AT.; de Andrade, M.; Doheny, KF.; Haines, JL.; Hayes, G., et al. Quality control procedures for genome-wide association studies. In: Haines, Jonathan L., et al., editors. Current protocols in human genetics / editorial board. Vol. Chapter 1. 2011. p. 19
27. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011; 478:103–109. [PubMed: 21909115]
28. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010; 42:348–354. [PubMed: 20208533]
29. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26:2336–2337. [PubMed: 20634204]
30. Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, Pistis G, Serbanovic-Canic J, Elling U, Goodall AH, Labrune Y, et al. New gene functions in megakaryopoiesis and platelet formation. Nature. 2011
31. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic epidemiology. 2010; 34:816–834. [PubMed: 21058334]
32. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, Chai HS, Bastarache L, Zuvich R, Peissig P, et al. Variants Near FOXE1 Are Associated with Hypothyroidism and Other Thyroid Conditions: Using Electronic Medical Records for Genome- and Phenome-wide Studies. Am J Hum Genet. 2011; 89:529–542. [PubMed: 21981779]
33. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, Chai HS, Bastarache L, Zuvich R, Peissig P, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. American journal of human genetics. 2011; 89:529–542. [PubMed: 21981779]
34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81:559–575. [PubMed: 17701901]
35. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. Genome research. 2003; 13:2129–2141. [PubMed: 12952881]
36. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, et al. Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res. 2009; 37:D619–D622. [PubMed: 18981052]
37. Du P, Feng G, Flatow J, Song J, Holko M, Kibbe WA, Lin SM. From disease ontology to disease-ontology lite: statistical methods to adapt a general-purpose ontology for the test of gene-ontology associations. Bioinformatics. 2009; 25:i63–i68. [PubMed: 19478018]
38. Watkins NA, Gusnanto A, de Bono B, De S, Miranda-Saavedra D, Hardie DL, Angenent WG, Attwood AP, Ellis PD, Erber W, et al. A HaemAtlas: characterizing gene expression in differentiated human blood cells. Blood. 2009; 113:e1–e9. [PubMed: 19228925]
39. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet. 2008; 40:395–402. [PubMed: 18311140]
40. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet. 2007; 39:857–864. [PubMed: 17554260]
41. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, et al. Genome-wide association study of blood pressure and hypertension. Nature genetics. 2009; 41:677–687. [PubMed: 19430479]

42. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, Thorleifsson G, Helgadottir H, Steinthorsdottir V, Stefansson H, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. Nature genetics. 2009; 41:342–347. [PubMed: 19198610]
43. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, Nakamura Y, Kamatani N. Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet. 2010; 42:210–215. [PubMed: 20139978]
44. Soranzo N, Spector TD, Mangino M, Kuhnel B, Rendon A, Teumer A, Willenborg C, Wright B, Chen L, Li M, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet. 2009; 41:1182–1190. [PubMed: 19820697]
45. Ferreira MA, Hottenga JJ, Warrington NM, Medland SE, Willemsen G, Lawrence RW, Gordon S, de Geus EJ, Henders AK, Smit JH, et al. Sequence variants in three loci influence monocyte counts and erythrocyte volume. Am J Hum Genet. 2009; 85:745–749. [PubMed: 19853236]
46. Meisinger C, Prokisch H, Gieger C, Soranzo N, Mehta D, Rosskopf D, Lichtner P, Klopp N, Stephens J, Watkins NA, et al. A genome-wide association study identifies three loci associated with mean platelet volume. Am J Hum Genet. 2009; 84:66–71. [PubMed: 19110211]
47. Soranzo N, Rendon A, Gieger C, Jones CI, Watkins NA, Menzel S, Doring A, Stephens J, Prokisch H, Erber W, et al. A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function. Blood. 2009; 113:3831–3837. [PubMed: 19221038]
48. Paul DS, Nisbet JP, Yang TP, Meacham S, Rendon A, Hautaviita K, Tallila J, White J, Tijssen MR, Sivapalaratnam S, et al. Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits. PLoS Genet. 2011; 7:e1002139. [PubMed: 21738486]
49. Qayyum R, Snively BM, Ziv E, Nalls MA, Liu Y, Tang W, Yanek LR, Lange L, Evans MK, Ganesh S, et al. A meta-analysis and genome-wide association study of platelet count and mean platelet volume in african americans. PLoS genetics. 2012; 8:e1002491. [PubMed: 22423221]
50. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, Oppermann U, Dilthey A, Pirinen M, Stone MA, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nature genetics. 2011; 43:761–767. [PubMed: 21743469]
51. Eriksson N, Tung JY, Kiefer AK, Hinds DA, Francke U, Mountain JL, Do CB. Novel associations for hypothyroidism include known autoimmune risk Loci. PLoS One. 2012; 7:e34442. [PubMed: 22493691]
52. Raychaudhuri S. Mapping rare and common causal alleles for complex human diseases. Cell. 2011; 147:57–69. [PubMed: 21962507]
53. Blattmann P, Schuberth C, Pepperkok R, Runz H. RNAi-based functional profiling of loci from blood lipid genome-wide association studies identifies genes with cholesterol-regulatory function. PLoS genetics. 2013; 9:e1003338. [PubMed: 23468663]
54. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, Raychaudhuri S. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nature genetics. 2013; 45:124–130. [PubMed: 23263488]
55. Kisacik B, Tufan A, Kalyoncu U, Karadag O, Akdogan A, Ozturk MA, Kiraz S, Ertenli I, Calguneri M. Mean platelet volume (MPV) as an inflammatory marker in ankylosing spondylitis and rheumatoid arthritis. Joint Bone Spine. 2008; 75:291–294. [PubMed: 18403245]
56. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola M, Jula A, Sinisalo J, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet. 2010; 376:1393–1400. [PubMed: 20971364]
57. Saade S, Cazier JB, Ghassibe-Sabbagh M, Youhanna S, Badro DA, Kamatani Y, Hager J, Yeretzian JS, El-Khazen G, Haber M, et al. Large scale association analysis identifies three susceptibility loci for coronary artery disease. PLoS One. 2011; 6:e29427. [PubMed: 22216278]
58. Hinks A, Martin P, Flynn E, Eyre S, Packham J, Barton A, Worthington J, Thomson W. Investigation of type 1 diabetes and coeliac disease susceptibility loci for association with juvenile idiopathic arthritis. Ann Rheum Dis. 2010; 69:2169–2172. [PubMed: 20647273]

59. Alcina A, Vandenbroeck K, Otaegui D, Saiz A, Gonzalez JR, Fernandez O, Cavanillas ML, Cenit MC, Arroyo R, Alloza I, et al. The autoimmune disease-associated KIF5A, CD226 and SH2B3 gene variants confer susceptibility for multiple sclerosis. Genes Immun. 2010; 11:439–445. [PubMed: 20508602]
60. Hunt KA, Zhernakova A, Turner G, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nature genetics. 2008; 40:395–402. [PubMed: 18311140]
61. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, Najjar SS, Zhao JH, Heath SC, Eyheramendy S, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nature genetics. 2009; 41:666–676. [PubMed: 19430483]
62. Tong W, Zhang J, Lodish HF. Lnk inhibits erythropoiesis and Epo-dependent JAK2 activation and downstream signaling pathways. Blood. 2005; 105:4604–4612. [PubMed: 15705783]
63. Wang DS, Shaw G. The association of the C-terminal region of beta I sigma II spectrin to brain membranes is mediated by a PH domain, does not require membrane proteins, and coincides with a inositol-1,4,5 triphosphate binding site. Biochemical and biophysical research communications. 1995; 217:608–615. [PubMed: 7503742]
64. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome research. 2012; 22:1748–1759. [PubMed: 22955986]
65. Movat HZ, Weiser WJ, Glynn MF, Mustard JF. Platelet phagocytosis and aggregation. J Cell Biol. 1965; 27:531–543. [PubMed: 4957257]
66. Kaplan JE, Saba TM. Platelet removal from the circulation by the liver and spleen. Am J Physiol. 1978; 235:H314–H320. [PubMed: 696842]
67. Italiano JE Jr, Lecine P, Shivdasani RA, Hartwig JH. Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes. J Cell Biol. 1999; 147:1299–1312. [PubMed: 10601342]
68. Denzer K, Kleijmeer MJ, Heijnen HF, Stoorvogel W, Geuze HJ. Exosome: from internal vesicle of the multivesicular body to intercellular signaling device. J Cell Sci. 2000; 113(Pt 19):3365–3374. [PubMed: 10984428]
69. Flaumenhaft R. Molecular basis of platelet granule secretion. Arterioscler Thromb Vasc Biol. 2003; 23:1152–1160. [PubMed: 12738684]
70. Reed GL. Platelet secretory mechanisms. Semin Thromb Hemost. 2004; 30:441–450. [PubMed: 15354265]
71. Patel SR, Hartwig JH, Italiano JE Jr. The biogenesis of platelets from megakaryocyte proplatelets. J Clin Invest. 2005; 115:3348–3354. [PubMed: 16322779]
72. Richardson JL, Shivdasani RA, Boers C, Hartwig JH, Italiano JE Jr. Mechanisms of organelle transport and capture along proplatelets during platelet production. Blood. 2005; 106:4066–4075. [PubMed: 16118320]
73. Suehiro Y, Veljkovic DK, Fuller N, Motomura Y, Masse JM, Cramer EM, Hayward CP. Endocytosis and storage of plasma factor V by human megakaryocytes. Thromb Haemost. 2005; 94:585–592. [PubMed: 16268476]
74. Blair P, Flaumenhaft R. Platelet-alpha-granules: basic biology and clinical correlates. Blood Rev. 2009; 23:177–189. [PubMed: 19450911]
75. Thon JN, Montalvo A, Patel-Hett S, Devine MT, Richardson JL, Ehrlicher A, Larson MK, Hoffmeister K, Hartwig JH, Italiano JE Jr. Cytoskeletal mechanics of proplatelet maturation and platelet release. J Cell Biol. 2010; 191:861–874. [PubMed: 21079248]
76. Fitzgerald DJ. Fibrinogen receptor and platelet signalling. Blood Coagul Fibrinolysis. 1999; 10(Suppl 1):S77–S79. [PubMed: 10070825]
77. Dhar A, Shukla SD. Tyrosine kinases in platelet signalling. Br J Haematol. 1993; 84:1–7. [PubMed: 8338765]
78. Gibbins JM. Platelet adhesion signalling and the regulation of thrombus formation. J Cell Sci. 2004; 117:3415–3425. [PubMed: 15252124]
79. Levy-Toledano S. Platelet signal transduction pathways: could we organize them into a 'hierarchy'? Haemostasis. 1999; 29:4–15. [PubMed: 10494030]

80. Li Z, Delaney MK, O'Brien KA, Du X. Signaling during platelet adhesion and activation. Arterioscler Thromb Vasc Biol. 2010; 30:2341–2349. [PubMed: 21071698]

81. Schulze H, Korpal M, Bergmeier W, Italiano JE Jr, Wahl SM, Shivdasani RA. Interactions between the megakaryocyte/platelet-specific beta1 tubulin and the secretory leukocyte protease inhibitor SLPI suggest a role for regulated proteolysis in platelet functions. Blood. 2004; 104:3949–3957. [PubMed: 15315966]

82. Qiao JL, Shen Y, Gardiner EE, Andrews RK. Proteolysis of platelet receptors in humans and other species. Biol Chem. 2010; 391:893–900. [PubMed: 20482312]

83. Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. Nature. 2010; 466:68–76. [PubMed: 20562859]

84. Wahlberg K, Jiang J, Rooks H, Jawaid K, Matsuda F, Yamaguchi M, Lathrop M, Thein SL, Best S. The HBS1L-MYB intergenic interval associated with elevated HbF levels shows characteristics of a distal regulatory region in erythroid cells. Blood. 2009; 114:1254–1262. [PubMed: 19528534]

85. Menzel S, Jiang J, Silver N, Gallagher J, Cunningham J, Surdulescu G, Lathrop M, Farrall M, Spector TD, Thein SL. The HBS1L-MYB intergenic region on chromosome 6q23.3 influences erythrocyte, platelet, and monocyte counts in humans. Blood. 2007; 110:3624–3626. [PubMed: 17712044]

86. Paul DS, Nisbet JP, Yang TP, Meacham S, Rendon A, Hautaviita K, Tallila J, White J, Tijssen MR, Sivapalaratnam S, et al. Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits. PLoS Genet. 2011; 7:e1002139. [PubMed: 21738486]

87. Ritchie MD, Denny JC, Zuvich RL, Crawford DC, Schildcrout JS, Bastarache L, Ramirez AH, Mosley JD, Pulley JM, Basford MA, et al. Genome- and phenome-wide analyses of cardiac conduction identifies markers of arrhythmia risk. Circulation. 2013; 127:1377–1385. [PubMed: 23463857]

88. Hebbring SJ, Schrodi SJ, Ye Z, Zhou Z, Page D, Brilliant MH. A PheWAS approach in studying HLA-DRB1*1501. Genes Immun. 2013; 14:187–191. [PubMed: 23392276]

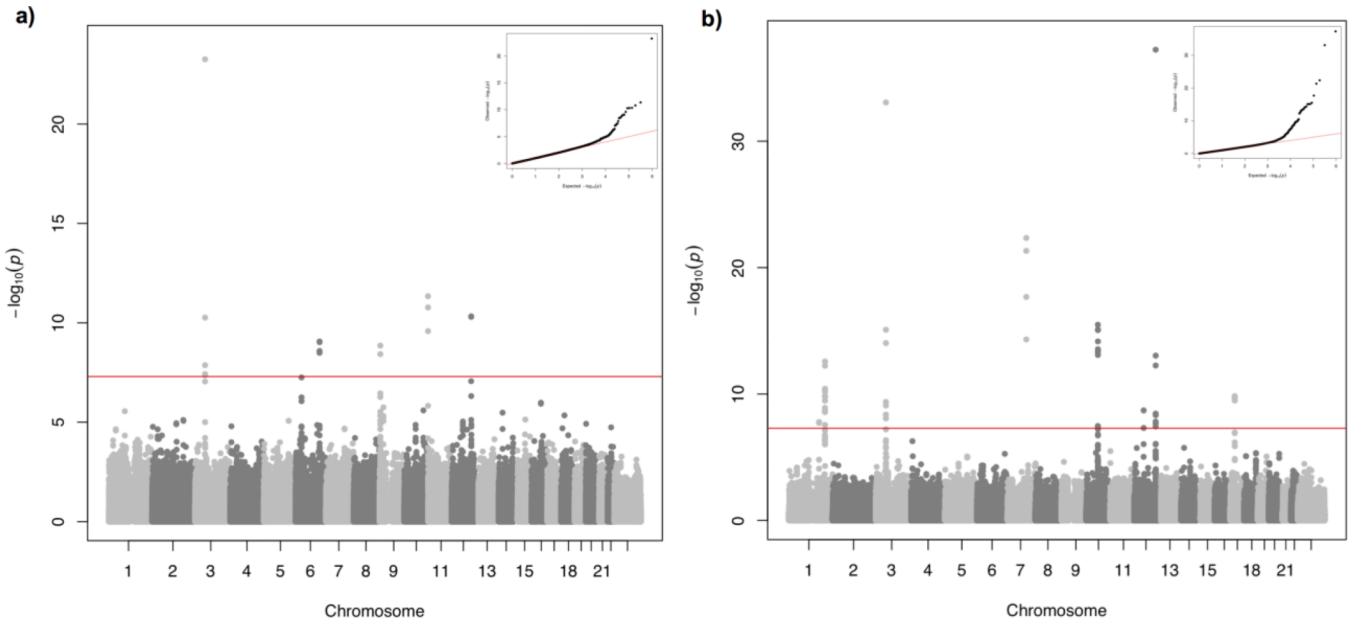

Figure 1. Manhattan and Quantile-Quantile Plots for Genome-wide Association Analysis of a) PLT (*n*=13, 852) and b) MPV(n=6, 291) in the eMERGE cohort

The vertical axis indicates (−log10 transformed) observed *P* values, and the horizontal line indicates the genome-wide significance level of p = 5E-8. In the quantile-quantile (QQ) plot, the horizontal axis shows (−log10 transformed) expected p values, and the vertical axis indicates (−log10 transformed) observed p values.

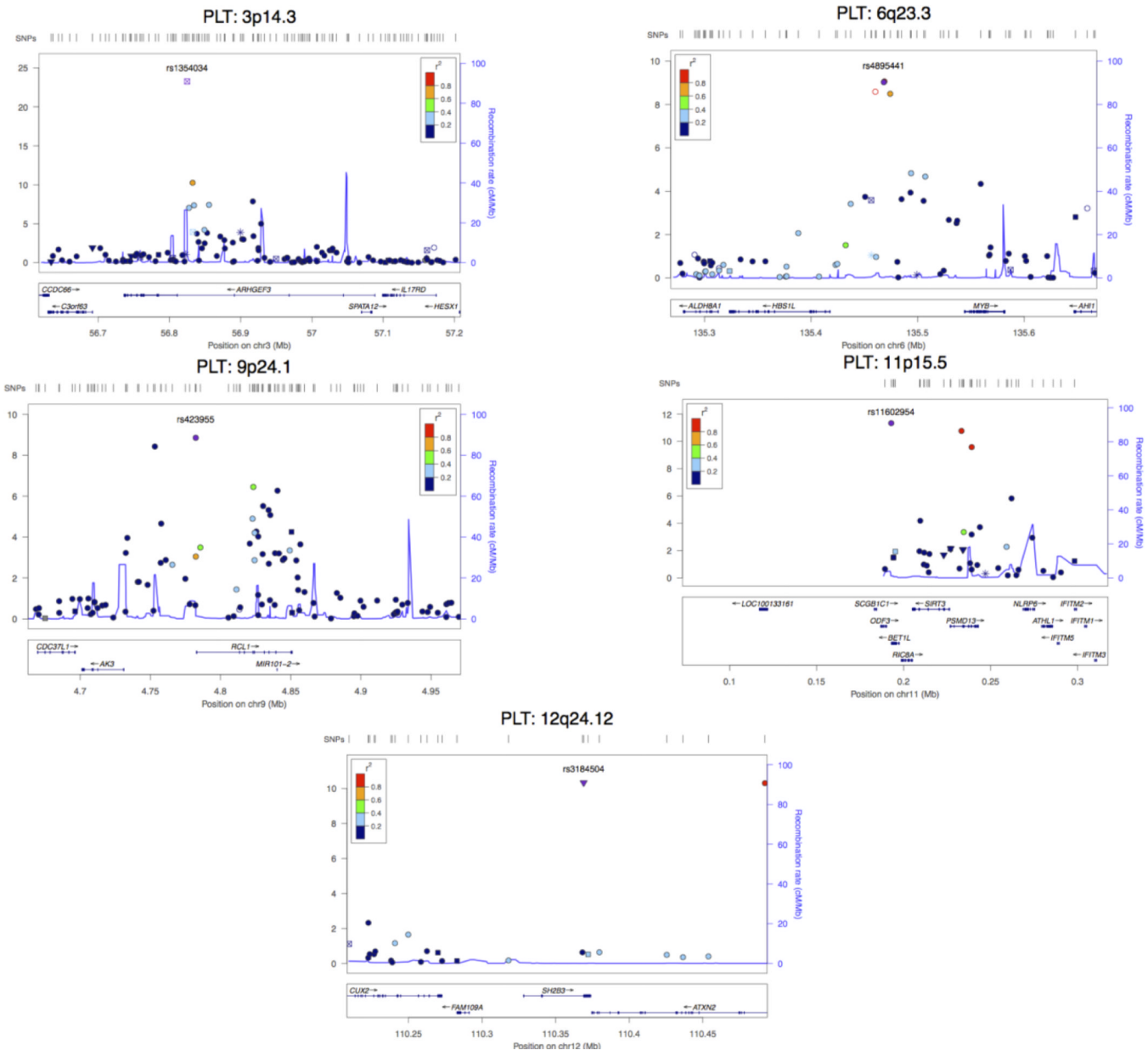

Figure 2. Regional plots of 5 loci associated with platelet count (PLT) on chromosomal regions: 3p14.3, 6q23.3, 9p24.1, 11p15.5 and 12q24.12

Single-nucleotide polymorphisms (SNPs) are plotted by position on the chromosome (x-axis) and the association with PLT ($-\log_{10} P$ value) on the y-axis. The rs number for the most significant SNP in the joint analysis is shown on the plot. Estimated recombination rates (from HapMap) are plotted in cyan to reflect the local linkage disequilibrium (LD) structure. The SNPs near the most significant SNP are color coded to depict their LD with the lead SNP (derived using pairwise r^2 values from the HapMap CEU data). Genes, the position of exons and the direction of transcription from the University of California, Santa Cruz genome browser are also plotted. ▾ = nonsynonymous; ○ = no annotation; ☒ = conserved in mammals; * = conserved TFBS (transcription factor binding site). cM/Mb = centimorgan/megabase; Mb = megabase.

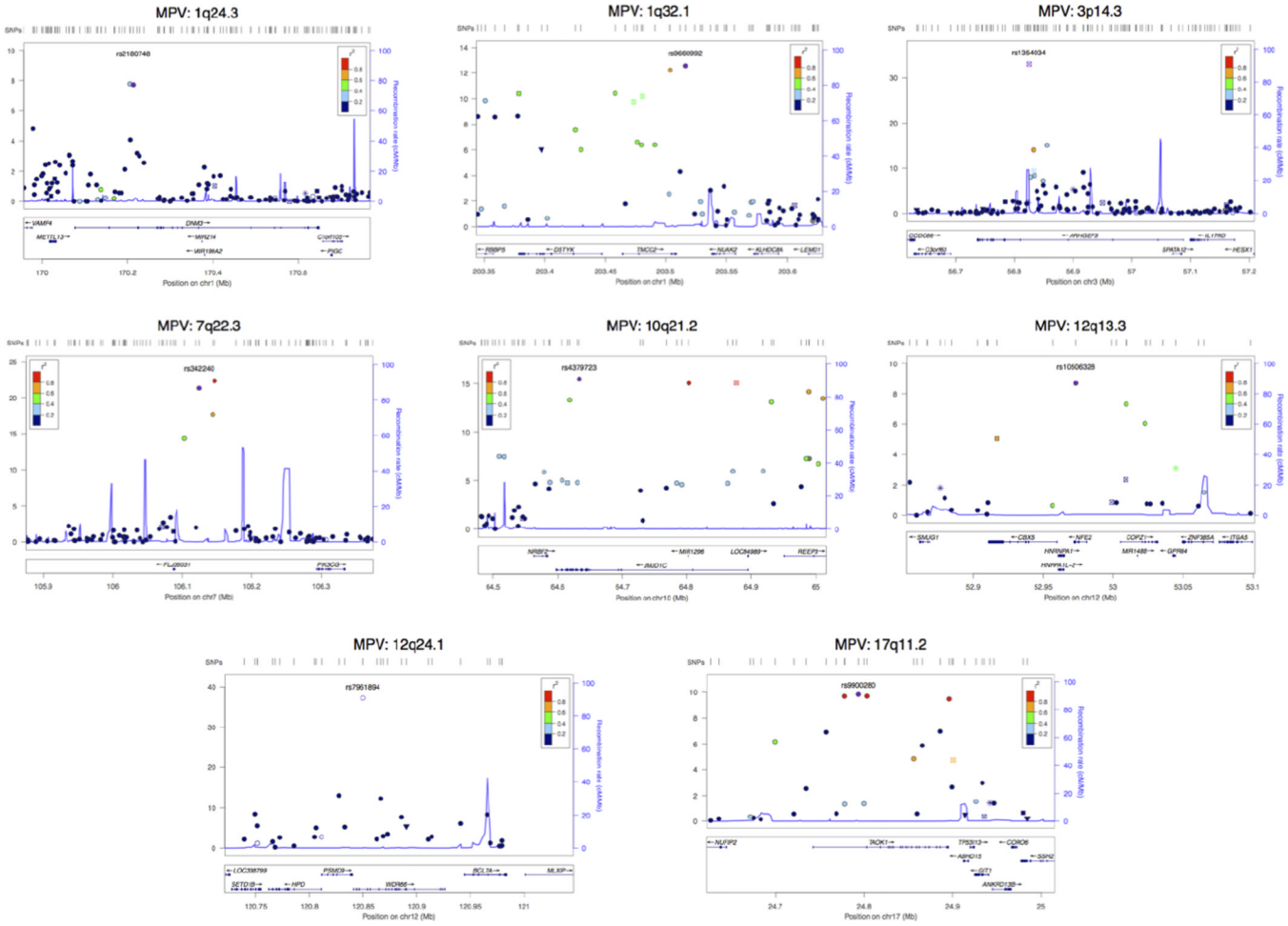

Figure 3. Regional plots of 8 loci associated with mean platelet volume (MPV) in chromosomal regions: 1q24.3, 1q32.1, 3p14.3, 7q22.3, 10q21.2, 12q13.3, 12q24.1 and 17q11.2

Single-nucleotide polymorphisms (SNPs) are plotted by position on the chromosome (x-axis) and the association with MPV ($-\log_{10}$ *P* value) on the y-axis. The rs number for the most significant SNP in the joint analysis is shown on the plot. Estimated recombination rates (from HapMap) are plotted in cyan to reflect the local linkage disequilibrium (LD) structure. The SNPs near the most significant SNP are color coded to depict their LD with the lead SNP (derived using pairwise r^2 values from the HapMap CEU data). Genes, the position of exons and the direction of transcription from the University of California, Santa Cruz genome browser are also plotted. ▾ = nonsynonymous; ○ = no annotation; ☒ = conserved in mammals; * = conserved TFBS (transcription factor binding site). cM/Mb = centimorgan/megabase; Mb = megabase.

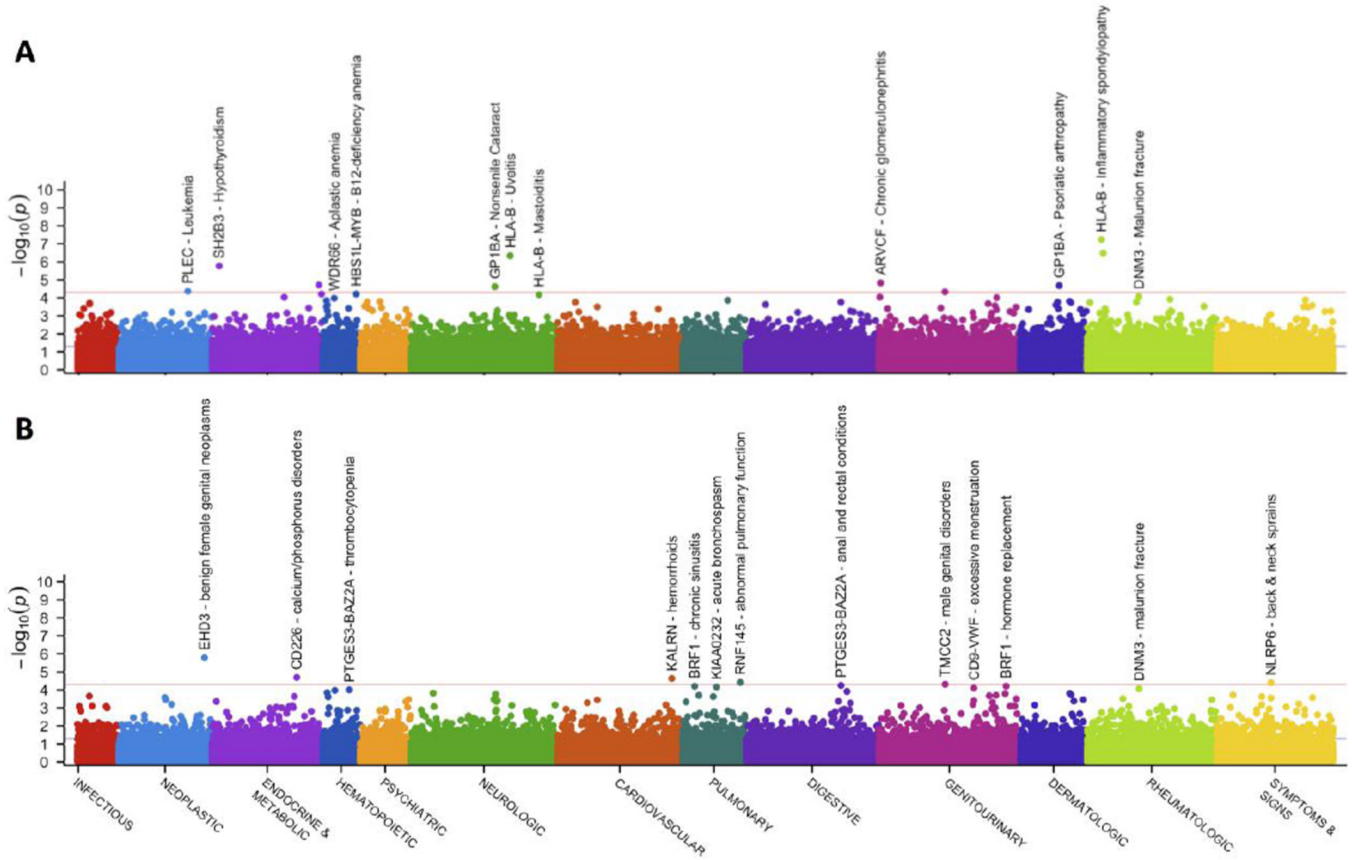

Figure 4. Manhattan plots summarizing results of a phenome-wide association analyses of A) platelet count and B) mean platelet volume

Results are from logistic regression analysis for 1368 EMR-defined phenotypes in 13,688 individuals, adjusted for age and gender. Broad phenotypic categories (X-axis) and −log10 (P-values) (Y-axis) are provided. Horizontal line indicates threshold for statistical significance (*P*-value 3.6E-5).

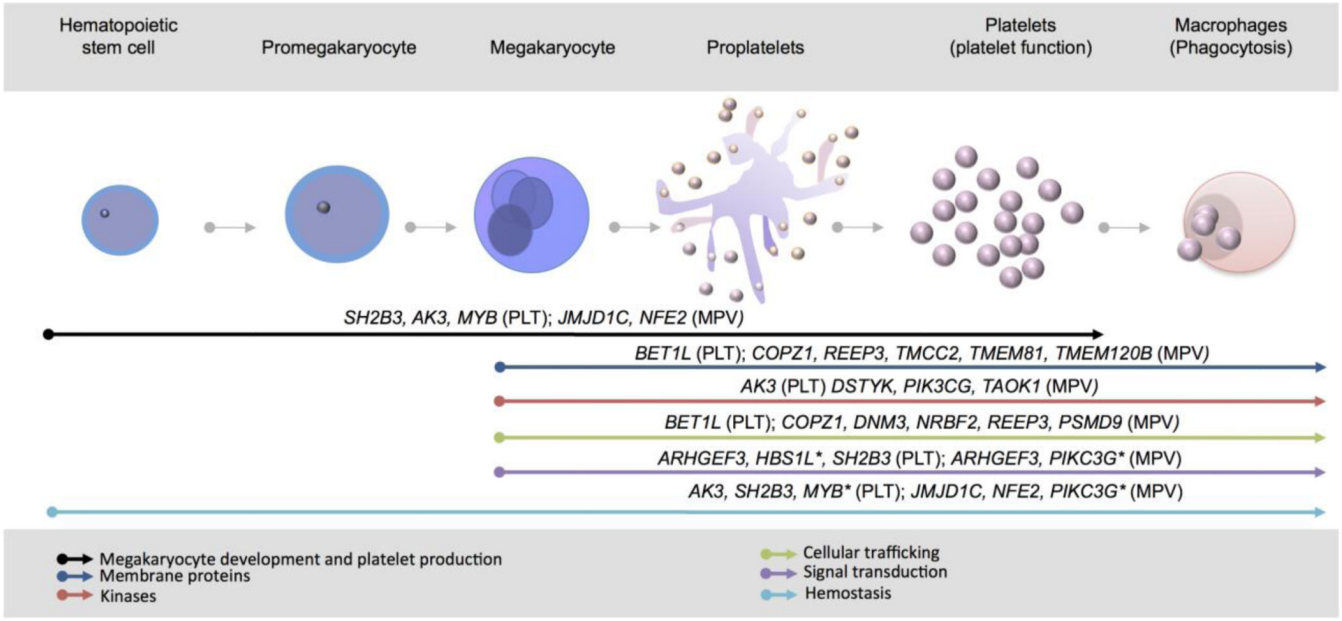

Figure 5. Functional repertoire of genes associated with PLT and MPV mapped to the platelet life cycle

Genes associated with variants influencing PLT and MPV are mapped to various stages of platelet biogenesis.

Table 1

Patient characteristics by site

	MC[1]	VUMC[2]	GHC[3]	MAYO[4]	NU[5]	Total
PLT (*n*)	3992	2923	2395	3187	1085	13,582
Men	1651 (41.3%)	1255 (42.9%)	1034 (43.7%)	1992 (62.5%)	518 (47.7%)	6450 (47.4%)
Age years	65.49 12.32	52.64 16.21	79.28 7.07	62.69 9.47	54.33 13.24	63.61 14.91
PLT*	247.35 59.25	261.18 71.43	247.65 67.94	241.04 61.11	249.99 66.15	249.11 64.93
MPV (*n*)	4013	NA	NA	2154	124	6,291
Men	1663 (41.4%)	NA	NA	1313 (60.9%)	51 (41.1%)	3027 (48.1%)
Age	66.00 12.31	NA	NA	59.73 9.19	56.60 13.25	63.67 11.78
MPV#	7.87 0.84	NA	NA	8.14 0.81	8.10 0.85	7.97 0.84

*10^3 /cumm

#femtolitre

1MC: Marshfield Clinic

2VUMC= Vanderbilt University Medical Centre

3GHC: Group Health Corporation

4Mayo: Mayo Clinic

5NU: Northwestern University

NA: Not available

Table 2

SNPs associated with platelet count (PLT) and mean platelet volume (MPV)

Trait	Lead SNP	Chr	Position[1]	Alleles[2]	Gene	Nearby genes	Feature	MAF[3]	beta[4]	*P*
PLT										
1	rs1354034	3p14.3	56824789	C/T	*ARHGEF3*	-	Intron	0.39	7.97	6E-24
2	rs4895441	6q23.3	135468266	A/G	-	*HBS1L-MYB*	Intergenic	0.27	−5.42	9E-10
3	rs423955	9p24.1	4782339	A/G	*RCL1*	-	5'reg	0.34	4.94	1E-09
4	rs11602954	11p15.5	192856	G/A	*BET1L*	-	3'reg	0.22	−6.46	5E-12
5	rs3184504	12q24.12	110368991	C/T	*SH2B3*	*ATXN2*	missense	0.50	−5.33	5E-11
MPV										
1	rs2180748	1q24.3	170213641	G/T	*DNM3*	-	intron	0.42	0.09	2E-08
2	rs9660992	1q32.1	203516073	A/G	*TMCC2*	*CNTN2, RBBP5, RIPK5,*	nearGene-3	0.43	0.11	3E-13
3	rs1354034	3p14.3	56824789	T/C	*ARHGEF3*	-	intron	0.39	−0.19	9E-34
4	rs342240	7q22.3	106124486	G/A	-	*FLJ36031-PIK3CG*	intergenic	0.44	−0.15	5E-22
5	rs4379723	10q21.2	64633455	T/C	*JMJD1C*	*NRBF2, REEP3*	intron	0.50	0.13	3E-16
6	rs10506328	12q13.3	52973499	C/A	*NFE2*	*COPZ1*	intron	0.36	−0.09	2E-09
7	rs7961894	12q24.1	120849966	C/T	*WDR66*	*TMEM120B, SETD1B, PSMD9, BCL7A*	intron	0.11	−0.31	6E-38
8	rs9900280	17q11.2	24793724	A/G	*TAOK1*	-	intron	0.48	0.1	1E-10

[1] NCBI 36

[2] The first allele is the major allele

[3] MAF, minor allele frequency

[4] The beta is the effect of the major allele

Table 3

Phenotypes associated by PheWAS PLT and MPV associated genetic variants

SNP	Disease[1]	Cases	Odds ratio	P[2]	Associated Phenotype	Nearest Gene
rs3819299	Inflammatory spondylopathies*	91	2.99	5.7E-08	PLT	*HLA-B*
rs3819299	Ankylosing spondylitis	23	5.45	3.3E-07	PLT	*HLA-B*
rs3819299	Uveitis (noninfectious)	261	2.10	4.6E-07	PLT	*HLA-B*
rs649729	Benign female genital neoplasms	22	4.76	1.7E-06	MPV	*EHD3*
rs3184504	Hypothyroidism	2120	1.18	1.7E-06	PLT	*SH2B3*
rs1034566	Chronic glomerulonephritis	32	2.97	1.5E-05	PLT	*ARVCF*
rs12969657	Calcium/phosphorus disorders	509	1.31	2.1E-05	MPV	*CD226*
rs6065	Psoriatic arthropathy	48	3.00	2.1E-05	PLT	*GP1BA*
rs10512627	Hemorrhoids	2912	0.87	2.3E-05	MPV	*KALRN*
rs6065	Nonsenile Cataract	238	1.88	2.4E-05	PLT	*GP1BA*
rs10076782	Abnormal pulmonary function studies	32	2.80	3.9E-05	MPV	*RNF145*
rs17655730	Sprains and strains of back and neck	2236	0.84	4.1E-05	MPV	*PSMD13–NLRP6*
rs6995402	Leukemia	204	0.64	4.3E-05	PLT	*PLEC*
rs1668871	Other disorders of male genital organs	313	1.42	4.7E-05	PLT	*TMCC2*
rs1172130	Other disorders of male genital organs	313	1.42	4.9E-05	MPV	*TMCC2*
rs2950390	Anal and rectal conditions	1033	0.82	5.6E-05	MPV	*PTGES3–BAZ2A*
rs9399137	B12-deficiency anemia	330	1.42	6.1E-05	PLT	*HBS1L–MYB*
rs3000073	Postmenopausal hormone replacement	584	0.75	6.1E-05	MPV	*BRF1*
rs6065	Other immunological findings	210	1.80	6.2E-05	PLT	*GP1BA*
rs3000073	Chronic sinusitis	2057	0.85	6.2E-05	MPV	*BRF1*
rs3819299	Mastoiditis & related conditions	28	3.79	6.7E-05	PLT	*HLA-B*
rs11734132	Acute bronchospasm	31	2.84	6.8E-05	MPV	*KIAA0232*
rs1558324	Excessive or frequent menstruation	466	1.32	7.9E-05	MPV	*CD9–VWF*
rs10914144	Malunion and nonunion of fracture	140	1.68	8.7E-05	MPV/PLT	*DNM3*
rs1034566	Glomerulonephritis	69	1.96	9.2E-05	PLT	*ARVCF*
rs7342306	Excessive or frequent menstruation	466	1.32	9.9E-05	PLT	*CD9–VWF*
rs2950390	Thrombocytopenia	628	0.78	9.9E-05	MPV	*PTGES3–BAZ2A*
rs7961894	Aplastic anemia	179	1.74	1.1E-04	MPV/PLT	*WDR66*

SNP	Disease[1]	Cases	Odds ratio	p[2]	Associated Phenotype	Nearest Gene
rs7317038	Gastrointestinal complications	283	1.39	1.2E-04	MPV	*GRTP1*
rs1260326	Osteoporosis, osteopenia, & pathological fractures	3352	1.13	1.2E-04	PLT	*GCKR*
rs6141	Early complications of trauma or procedure	68	0.49	1.3E-04	PLT	*THPO*
rs559972	Respiratory failure; insufficiency; arrest	539	0.78	1.4E-04	PLT	*TAOK1*
rs10761731	Megaloblastic anemia	185	1.50	1.4E-04	PLT	*JMJD1C*
rs7075195	Megaloblastic anemia	185	1.50	1.4E-04	MPV	*JMJD1C*
rs3000073	Headache syndromes	2304	0.87	1.5E-04	MPV	*BRF1*
rs1354034	Nail diseases	1067	1.20	1.6E-04	MPV/PLT	*ARHGEF3*
rs505404	Mood disorders	2891	0.87	1.6E-04	PLT	*PSMD13*
rs3000073	Symptomatic artificial menopause	153	0.58	1.6E-04	MPV	*BRF1*
rs11082304	Other disorders of bone and cartilage	465	0.78	1.7E-04	PLT	*CABLES1*
rs6065	Psoriasis	345	1.59	1.7E-04	PLT	*GP1BA*
rs3184504	Myocardial infarction	1400	1.18	1.7E-04	PLT	*SH2B3*
rs10512627	Refraction and accommodation disorders	6082	0.90	1.8E-04	MPV	*KALRN*
rs7149242	Dermatomyositis and Polymyositis	34	2.51	1.8E-04	PLT	*C14orf70 – DLK1*
rs7616006	Streptococcus infection	187	0.66	2.0E-04	PLT	*SYN2*
rs1558324	Irregular menstrual cycle/bleeding	1009	1.22	2.0E-04	MPV	*CD9–VWF*
rs3000073	Other upper respiratory disease	1144	0.83	2.0E-04	MPV	*BRF1*
rs10761731	Hypertrophy of female genital organs	57	2.03	2.1E-04	PLT	*JMJD1C*
rs7075195	Hypertrophy of female genital organs	57	2.03	2.1E-04	MPV	*JMJD1C*
rs8109288	Streptococcus infection	187	4.93	2.2E-04	MPV/PLT	*TPM4*
rs12969657	Disorders of mineral metabolism	960	1.19	2.3E-04	MPV	*CD226*

*Includes ankylosing spondylitis

[1]Results include the top 50 associated phenotypes from logistic regression adjusted for age and gender

[2]A Bonferroni correction for PheWAS performed on a single SNP is 0.05/1368=3.6E-5