

Genome sequencing unveils a regulatory landscape of platelet reactivity

Ali R. Keramati^{1,2,124}, Ming-Huei Chen^{3,4,124}, Benjamin A. T. Rodriguez^{3,4,5,124}, Lisa R. Yanek^{2,6}, Arunoday Bhan⁷, Brady J. Gaynor^{8,9}, Kathleen Ryan^{8,9}, Jennifer A. Brody¹⁰, Xue Zhong¹¹, Qiang Wei¹², NHLBI Trans-Omics for Precision (TOPMed) Consortium*, Kai Kammers¹³, Kanika Kanchan¹⁴, Kruthika Iyer¹⁴, Madeline H. Kowalski¹⁵, Achilleas N. Pitsillides^{4,16}, L. Adrienne Cupples^{4,16}, Bingshan Li¹², Thorsten M. Schlaeger⁷, Alan R. Shuldiner⁹, Jeffrey R. O'Connell^{8,9}, Ingo Ruczinski¹⁷, Braxton D. Mitchell^{8,9}, Nauder Faraday^{2,18}, Margaret A. Taub¹⁷, Lewis C. Becker^{1,2}, Joshua P. Lewis^{8,9,125}✉, Rasika A. Mathias^{2,14,125}✉ & Andrew D. Johnson^{3,4,125}✉

Platelet aggregation at the site of atherosclerotic vascular injury is the underlying pathophysiology of myocardial infarction and stroke. To build upon prior GWAS, here we report on 16 loci identified through a whole genome sequencing (WGS) approach in 3,855 NHLBI Trans-Omics for Precision Medicine (TOPMed) participants deeply phenotyped for platelet aggregation. We identify the *RGS18* locus, which encodes a myeloerythroid lineage-specific regulator of G-protein signaling that co-localizes with expression quantitative trait loci (eQTL) signatures for *RGS18* expression in platelets. Gene-based approaches implicate the *SVEP1* gene, a known contributor of coronary artery disease risk. Sentinel variants at *RGS18* and *PEAR1* are associated with thrombosis risk and increased gastrointestinal bleeding risk, respectively. Our WGS findings add to previously identified GWAS loci, provide insights regarding the mechanism(s) by which genetics may influence cardiovascular disease risk, and underscore the importance of rare variant and regulatory approaches to identifying loci contributing to complex phenotypes.

¹ Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ² GeneSTAR Research Program, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ³ Division of Intramural Research, Population Sciences Branch, National Heart, Lung and Blood Institute, Bethesda, MD, USA. ⁴ The Framingham Heart Study, Framingham, MA, USA. ⁵ Valo Health, Boston, MA, USA. ⁶ Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁷ Boston Children's Hospital, Boston, MA, USA. ⁸ Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA. ⁹ Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. ¹⁰ Cardiovascular Health Research Unit, University of Washington School of Medicine, Seattle, WA, USA. ¹¹ Vanderbilt Genetics Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ¹² Vanderbilt Genetics Institute, Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA. ¹³ Biostatistics and Bioinformatics, Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ¹⁴ Division of Allergy and Clinical Immunology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ¹⁵ Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA. ¹⁶ Department of Biostatistics, School of Public Health, Boston University, Boston, MA, USA. ¹⁷ Bloomberg School of Public Health, Biostatistics, Johns Hopkins University, Baltimore, MD, USA. ¹⁸ Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ¹²⁴ These authors contributed equally: Ali R. Keramati, Ming-Huei Chen, Benjamin A.T. Rodriguez. ¹²⁵ These authors jointly supervised this work: Joshua P. Lewis, Rasika A. Mathias, Andrew D. Johnson. *A list of authors and their affiliations appears at the end of the paper. ✉email: jlewis2@som.umaryland.edu; rmathias@jhmi.edu; johnsonad2@nhlbi.nih.gov

Atherosclerotic cardiovascular diseases (ASCVD) have remained the major cause of morbidity and mortality worldwide. The hallmark of ASCVD is aggregation of activated platelets on a ruptured atherosclerotic plaque followed by thrombus formation¹. Hemostasis and platelet aggregation is an evolutionary conserved process that is maintained by a delicate balance between agonists like ADP and epinephrine and antagonists like prostaglandins². Prior studies have shown that platelet aggregation in response to agonists is highly heritable with heritability estimates between 40 and 60%^{3–5}. High platelet reactivity at baseline and after inhibition with aspirin is associated with poor cardiovascular outcome^{6,7}. Antiplatelet therapies are standard-of-care for secondary prevention of the complications of occlusions in coronary, cerebral, and peripheral arteries. Prior genome- and exome-wide association studies have identified at least 8 common variants for platelet aggregation in response to different agonists^{8–11}. With the exception of a few limited gene-based scans^{9,12}, no previous genome-wide studies have systematically evaluated the contribution of both common and rare variants to heritability of agonist-induced platelet reactivity. Thus, it is likely that significant missing heritability remains for platelet function traits.

In this work leveraging the scientific resources of the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program, we report the first association study of platelet aggregation in response to variety of physiological stimuli using whole-genome sequencing (WGS) data. We sought to 1) refine previously identified GWAS loci, 2) identify novel loci that determine platelet aggregation in response to different doses of ADP, epinephrine and collagen, 3) examine the collective burden of coding variants on platelet aggregation, and 4) evaluate the collective burden of rare non-coding variants of megakaryocyte-specific super-enhancer regions on platelet aggregation. Extension of genetic findings using biobank resources as well as ex vivo cell-based functional systems were also performed.

Results

Single-variant based tests for association. There were a total of 19 harmonized phenotypic measures of platelet aggregation evaluated in this investigation (Supplementary Table 1). This includes 9 phenotypes for adenosine diphosphate (ADP) as an agonist, 9 for epinephrine, and 4 for collagen. Genome-wide single variant tests for association were performed on ~28 million variants in 3,125 European Americans (EA) and 730 African Americans (AA) (Supplementary Table 2) from the Framingham Heart Study (FHS), Older Order Amish Study (OOA), and the Genetic Study of Atherosclerosis Risk (GeneSTAR). We identified 101 variants associated with platelet aggregation in response to ADP, epinephrine, or collagen (P value $< 5 \times 10^{-8}$, Fig. 1, Supplementary Fig. 1). Using iterative conditional analyses, genome-wide significant variants were refined down to 16 independent loci (Table 1). With the exception of two variants (rs12041331 and chr17:21960955) all loci were associated with platelet aggregation in response to a single agonist (Fig. 1B), and most of the identified loci were not present in the prior array-based approaches^{8–10} (Table 1, Supplementary Figs. 2–4).

Replication of the single-variant results. Replication of discovery findings was performed in up to 2,009 independent samples from FHS, OOA, and GeneSTAR (Supplementary Data 1), and extended into an independent cohort (the Caerphilly Prospective Study [CaPS], $N = 1183$) for ADP and collagen-induced platelet aggregation phenotypes^{8,13} (Supplementary Table 4). Among the 7 previously reported loci¹⁰, 2 were replicated in this investigation (*PEAR1* and *ADRA2A*, Table 1).

Reduction in sample size, a low overlapping percentage (<75%) of participants in 2 of the previously studied cohorts (FHS and GeneSTAR European samples), addition of subjects (OOA and GeneSTAR African Americans), and the difference between WGS data and HapMap imputed dosage data (Supplementary Table 5a) may explain, in part, the lack of association observed with the other 5 previously identified loci. Meta-analysis, as opposed to mega-analysis approaches, did not meaningfully change the interpretation of these findings (Supplementary Table 5a, b) comparing the current WGS results to prior studies. Comparison of previous results with the current investigation for the *RGS18* variant is shown in Supplementary Table 5b; all other newly-identified WGS variants from this study were not available in the previous investigation.

Co-localization of the genetic loci with eQTLs in platelets.

Given that all 16 loci identified using single-variant approaches are located in non-coding regions of the genome, we tested for co-localization between these regions and eQTL data available through RNA sequencing of platelets in 180 European Americans from GeneSTAR (Supplementary Table 6). We found that sentinel variants in the *PEAR1* and *RGS18* loci were eQTLs for *PEAR1* and *RGS18*, respectively. No co-localization was noted for any of the remaining 14 loci. As noted in Fig. 2a, there is likely only a single variant accounting for the *PEAR1* GWAS peak, in contrast to *RGS18* where there are likely several causal variants.

PheWAS in external Biobanks. An examination of the sentinel variants reported in Table 1 was performed in the UK Biobank and BioVU as presented in Supplementary Data 2. The minor allele (A) of *PEAR1* at rs12041331, which is known to be associated with lesser platelet aggregation, *PEAR1* RNA, and protein expression^{14,15}, was associated with increased odds of gastrointestinal bleeding in both EAs and AAs in the BioVU Biobank PheWAS.

Functional follow up of the *RGS18* locus. In the *RGS18* region, several variants were replicated using independent samples (Supplementary Data 1), and additional evidence was also observed for ADP and collagen aggregation phenotypes in the CaPS study (Supplementary Table 4). Overlaying the associated variants with platelet eQTLs and megakaryocytic epigenome features, there are several potential candidate polymorphisms (Supplementary Table 8). Consistent with our human results, independent *Rgs18*^{-/-} mouse studies suggest *Rgs18* inhibits pre-agonist stimulated platelet reactivity, with knockouts exhibiting exaggerated platelet reactivity to multiple agonist pathways, decreased bleeding times, and increased arterial occlusion^{16,17}. This is attributed to a loss of inhibition of multiple G-protein coupled receptor signaling pathways in platelets¹⁸. The minor allele (C) of *RGS18* at rs1175170, is associated with arterial thrombosis/embolization in both EAs and AAs in the BioVU Biobank (Supplementary Data 2). Allele-specific and transcription-factor overexpression studies suggest that rs12070423, which may disrupt a GATA1 target site, and rs4495675, which may disrupt a NFE2 target site, both reduce *RGS18* expression (Fig. 2b, Supplementary Table 9, Supplementary Figs. 4 and 5). These SNPs are in LD (minimum r^2 0.614) with rs1175170 suggesting they may be functional variants on the same haplotype that affect *RGS18*-mediated platelet activation.

Genes identified through rare variant based approaches. SKAT¹⁹ gene-based tests using a MAF threshold of 0.05 were

Table 1 Sixteen loci identified through single variant approaches for genome-wide association for platelet aggregation in response to epinephrine, ADP and collagen in 3855 TOPMed participants.

Known vs Novel ^a	chr:pos (hg38)	rsID	ref/alt	Nearest Gene	MAF	ADP		Collagen lag time		Epinephrine	
						P	beta	P	beta	P	beta
Novel	1:20567949	rs12137738	A/T	FAM43B,CDA	0.077	2.62E-04	0.22	1.71E-02	-0.108	1.04E-08	0.306
Novel	1:67128641	rs142001088	C/T	C1orf141	0.018	2.88E-02	0.203	9.25E-09	-0.503	2.51E-03	0.276
Known	1:156899922	rs12041331	G/A	PEAR1	0.148	7.61E-17	-0.329	7.58E-17	0.317	2.31E-18	-0.358
Novel	1:192194880	rs1175170	G/C	RGS18,RGS21	0.442	7.86E-06	0.123	2.37E-02	-0.057	1.96E-09	0.155
Novel	5:19109993	rs112157462	T/C	LINC02223, CDH18	0.022	1.64E-02	-0.281	1.19E-08	0.458	5.90E-03	-0.23
Novel	6:121921871	rs58250884	A/G	GJA1,HSF2	0.087	1.69E-03	-0.153	2.34E-02	0.106	2.22E-08	-0.273
Novel	9:28873884	rs185159562	T/A	LINGO2	0.005	1.16E-02	-0.447	1.45E-01	0.243	3.87E-08	-0.988
Novel	10:75490891	rs138028657	A/G	LRMDA	0.006	6.44E-01	0.073	1.52E-08	-0.858	6.14E-01	-0.102
Known	10:111139289	rs7097060	T/A	ADRA2A,GPAM	0.137	5.45E-01	-0.028	6.73E-01	-0.015	6.68E-12	-0.251
Novel	11:92185065	rs183146849	A/T	DISC1FP1,FAT3	0.012	3.11E-08	-0.702	4.38E-01	-0.084	5.73E-04	-0.376
Novel	12:132589485	rs140148392	G/A	FBRSL1,LRCOL1	0.009	1.51E-01	0.254	2.88E-08	0.669	3.54E-01	0.138
Novel	13:96912429	rs61974290	A/G	HS6ST3, LINC00359	0.057	9.58E-02	-0.088	9.43E-03	0.134	3.36E-09	-0.4
Novel	17:16451482	rs575524466	G/A	LRRCT5A-AS1	0.003	1.90E-01	-0.315	3.13E-08	1.169	5.35E-02	-0.463
Novel	17:21960955	-	A/T	KCNJ18,UBBP4	0.276	2.87E-08	0.167	1.88E-03	-0.089	7.73E-19	0.26
Novel	18:29059923	rs138845468	TAAATA/T	CDH2,MIR302F	0.082	6.99E-02	0.087	3.23E-08	-0.25	1.21E-01	0.087
Novel	20:50142397	rs542707094	CTG/C	TMEM189, TMEM189-UBE2V1	0.003	3.16E-02	-0.486	3.53E-08	1.194	3.05E-02	-0.487

P values presented are a summary across all individual phenotypes for the single agonist (i.e., the minimum P value for the SNV from 8, 7 and 4 individual phenotypes for epinephrine, ADP, and collagen, respectively), alternate allele (alt) represents minor allele (see Supplementary Table 1). P values are from a two-sided score test with no adjustment for multiple testing.

^aLoci were defined as known if they were identified in the prior array-based GWAS approaches⁸⁻¹⁰, else they were labeled as novel.

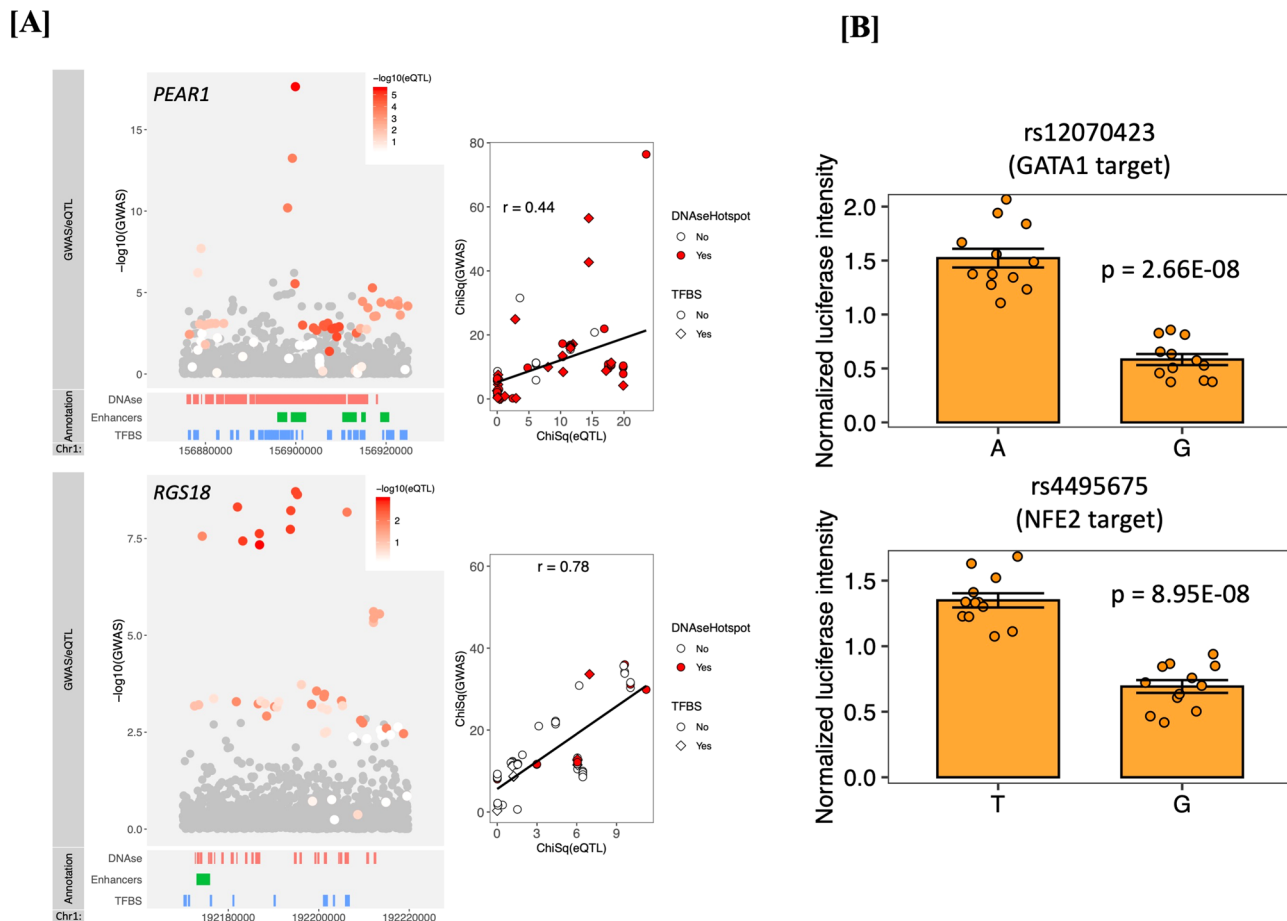


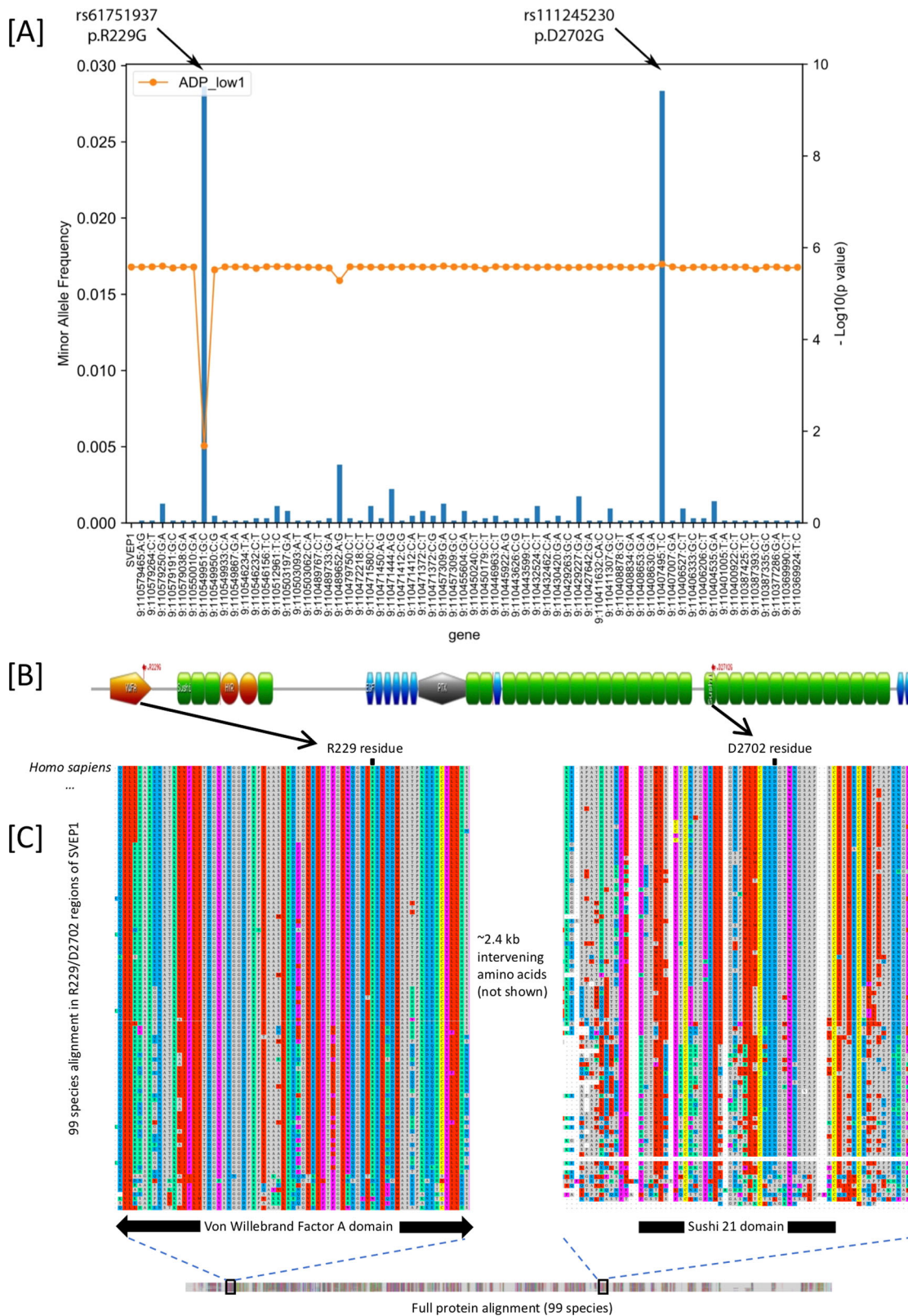
Fig. 2 Co-localization of WGS signals in *PEAR1* and *RGS18* regions with effects on platelet gene expression and regulatory features and demonstration of *RGS18* allele-specific SNP effects on enhancer activity. Co-localization of WGS association signal and platelet eQTL signatures and allele-specific experiments for *RGS18* SNP enhancers. *P* values in panel **A** are two-sided score/linear model tests for GWAS/eQTLs, respectively, with no adjustment for multiple testing. *P* values in Panel **B** are from a two-sided Welch test with no adjustment for multiple testing. **A** Top panels show co-localization between *PEAR1* eQTL and WGS association for platelet aggregation in response to Epi_low1 (see Supplementary Table 1) and bottom panels are between *RGS18* eQTL with WGS association for platelet aggregation in response to Epi_low 5 (see Supplementary Table 1). In the left panels, the region of co-localization is zoomed to the sentinel SNP ± 25 kb, the Y axis shows the $-\log(P)$ of the GWAS association, the color of the dot represents the strength of the eQTL evidence for the gene, and SNVs that were not included in eQTL analysis are shown in gray. The right panels show the scatter plot and correlation between the ChiSquare statistic for the GWAS and eQTL signal for all SNVs present in both sets of data. **B** Allele-specific enhancer activity differences for rs12070423 A or G alleles in HEK293 cells lentiviral transfected to overexpress GATA1 (top) and rs4495675 T or G alleles in HEK293 cells lentiviral transfected to overexpress NFE2 (bottom). Each allele-specific result represents results of 12 experiments (12 biological replicates over 3 independent replicates). Data presented represent mean values \pm SEM.

and thrombotic outcomes including occlusions, cerebrovascular disease, cardiac arrest, embolism and deep vein thrombosis suggest that *RGS18* may be a critical node for intervention in platelets.

The WGS approach allowing for a rare-variant gene-based analysis suggests that *SVEP1* may have previously unappreciated and multifactorial roles in contributing to CVD. Homozygous *Svep1*^{-/-} mice die from edema, and heterozygous mice, as well as zebrafish, experience arterial and lymphatic vessel malformations^{25–27}. Consistent with previous investigations²⁸, RNAseq data in a subset of GeneSTAR participants do not indicate expression of *SVEP1* in platelets. We find that rs61751937 is the strongest plasma protein QTL for *SVEP1* (*P* value = 5.2×10^{-64}), reducing expression²⁹, suggesting the effects may be mediated through interactions of platelets with other cell types in circulation. This conserved protein could potentially affect platelet function and CVD through several mechanisms including cell-cell adhesion, cell differentiation, and functions in bone marrow niches³⁰. Recent functional work demonstrates that

SVEP1 is expressed in plaques. Further experiments suggested that deficiency of *Svep1* affects Cxcl1 endothelial release and promotes proinflammatory leukocyte recruitment to plaques³¹. Given our assays are ex vivo assessments of platelet function in PRP lacking endothelial, leukocyte and smooth muscle cells, this suggests that alteration of *SVEP1* levels or other related factors in plasma may also have direct effects on platelets that may influence thrombus formation.

In conclusion, there is a large body of evidence supporting the hypothesis that hyper-reactive platelets may predict future thromboses in both healthy individuals⁶ and those who have already experienced thrombosis^{32,33}. Therefore, better understanding of the genetic determinants of heightened platelet aggregation is likely critical in the early prediction of thrombosis events as well as aiding in pharmacogenetic efforts pertaining to antiplatelet therapy. By applying contemporary WGS strategies in participants with extensive platelet reactivity phenotype data, we show the potential for such approaches to identify genetic determinants that may impact such traits.



Methods

Description of study populations

GeneSTAR. The Genetic Study of Atherosclerosis Risk (*GeneSTAR*) is an ongoing, prospective family-based study designed to explore environmental, phenotypic, and genetic causes of premature cardiovascular disease. Participants were recruited from European- and African-American families ($n = 891$) identified from probands who were hospitalized for a coronary disease event prior to 60 years of age in any of 10 Baltimore, Maryland area hospitals. Apparently healthy siblings of the

probands, offspring of the siblings and probands, and the co-parents of the offspring were screened for traditional coronary disease and stroke risk factors as part of a study of platelet function prior to and following a 2-week trial of 81 mg/day of aspirin from 2003 to 2006^{3,34}. All measures described here were obtained prior to the commencement of aspirin. Exclusion criteria included: 1) any coronary heart disease or vascular thrombotic event, 2) any bleeding disorder or hemorrhagic event (e.g., stroke or gastrointestinal bleed), 3) current use of any anticoagulants or antiplatelet agents (i.e., warfarin, persantin, clopidogrel), 4) current use of chronic

Fig. 3 Protein-coding variant effects in SVEP1 on ADP platelet activation, protein domains and cross-species conservation of 229G, 2702D and surrounding sequence. Association of aggregated rare deleterious coding variants in *SVEP1* and ADP-induced platelet aggregation. *P* values in panel **A** are from two-sided score tests with no adjustment for multiple testing. **A** Using a *leave-one-out* approach, we identified a rare coding variant (rs61751937) that explains most of the association. The first orange dot represents the overall gene-based test including the full set of 64 variants. Subsequent orange dots represent the $-\log_{10}(P)$ of the gene-based test when the specific labeled variant was left out, and blue bars represent minor allele frequency of specific variant being left out. **B** Schematic protein structure of *SVEP1*. rs61751937 substitutes glycine for arginine at position 2702. Another variant in *SVEP1* has been associated with coronary artery disease which substitutes glycine for aspartic acid at position 2702. **C** Using UniProt, a total of 98 orthologs were identified for the largest human *SVEP1* protein isoform and aligned. Alignments were visualized in MAFFT (v.7, <https://mafft.cbrc.jp/alignment/server/>, Katoh et al. 2017)⁵⁹ with ClustalW coloring. Both amino acids 229G and 2702D are highly conserved across diverse species, as well as their surrounding protein domains. The sequence identifiers and genus and species are given in Supplementary Data 3.

or acute nonsteroidal anti-inflammatory agents, including COX-2 inhibitors that could not be discontinued, 5) recent active gastrointestinal disorder, 6) current pharmacotherapy for a gastrointestinal disorder, 7) pregnancy or risk of pregnancy during the trial, 8) recent menorrhagia, 9) known aspirin intolerance or allergic side effects, 10) serious medical disorders, (e.g., autoimmune diseases, renal or hepatic failure, cancer or HIV-AIDS), 11) current chronic or acute use of glucocorticosteroid therapy or any drug that may interfere with the measured outcomes, 12) serious psychiatric disorders, and, 13) inability to independently make a decision to participate. Of the 3003 participants in the aspirin trial, 1786 were selected for whole-genome sequencing (WGS) in the Trans-Omics for Precision Medicine (TOPMed) Program based on 1) complete platelet function phenotyping and 2) largest family size.

Framingham Heart Study. The Framingham Heart Study (FHS) is a longitudinal family-based study that started to recruit participants of European ancestry in 1948 and now is on its third generation of participants. The Original cohort (first generation) contains 5209 participants, the Offspring cohort (second generation), began to recruit in 1971, contains 5124 participants, and the Third Generation cohort, began to recruit in 2002, contains 4095 participants. In the present study, we use data from the Offspring cohort¹⁰. For FHS, aspirin use was determined based on arachidonic acid and review of platelet aggregation curves.

Old Order Amish (OOA). As part of the Amish Complex Disease Research Program, a prospective cohort trial examining the relationship between genetic variants and agonist-induced platelet function at baseline and in response to clopidogrel and aspirin was performed. Characteristics of this cohort have been described previously³⁵. Briefly, Amish participants who were over age 20, generally healthy, and agreed to discontinue the use of medications, supplements, and vitamins for at least one week prior to study initiation were eligible for recruitment. Medical and family histories, anthropometry, physical examinations, and blood samples were obtained after an overnight fast. All measures described here were obtained prior to clopidogrel or aspirin administration. Participants were excluded from participation if any of the following criteria were met: 1) currently pregnant or breastfeeding, 2) history of a bleeding disorder or major spontaneous bleed, 3) severe hypertension (bp >160/95 mm Hg), 4) coexisting malignancy, 5) creatinine >2.0 mg/dl, 6) AST or ALT >2 times the upper limit of normal, 7) Hct <32%, 8) TSH <0.4 or >5.5 mIU/L, 9) platelet count >500,000/ul or <75,000/ul, 10) surgery within the last 6 months, 11) allergy to aspirin or clopidogrel, or 12) unwilling or unable to discontinue any medications that may interfere with the results of the study outcomes.

Written informed consent was obtained from all participants, and each study was approved by their local review board (GeneSTAR- Johns Hopkins Institutional Review Board; FHS- Boston University Institutional Review Board; and OOA- University of Maryland, Baltimore Institutional Review Board).

Platelet function tests and phenotype harmonization. Methods to assess ex vivo platelet function have been described in detail previously^{8,10}. In brief, blood samples were obtained after an overnight fast into 3.2% (or 3.8% in FHS) citrated vacutainer tubes. Platelet-rich and platelet-poor plasma (PRP and PPP, respectively) were isolated by centrifugation (PRP, 180 × g for 15 min in GeneSTAR and OOA, 160 × g for 5 min in FHS; PPP 2000 × g for 10 min in GeneSTAR and OOA, 2500 × g for 20 min for FHS). Light transmittance aggregometry was performed in PRP using a PAP-4 (GeneSTAR and FHS) or a PAP-8E (OOA) aggregometer after stimulation with ADP, epinephrine, or collagen using PPP as a referent. In GeneSTAR, maximal aggregation (% aggregation) was recorded for periods of 5 min after stimulation with ADP (2.0 and 10.0 μM, Chronolog Corp, Haverton, PA) or epinephrine (2.0 and 10.0 μM, Chronolog Corp, Haverton, PA); and lag time to initiation of aggregation was recorded after stimulation with equine tendon-derived type I collagen (1, 2, 5 and 10 μg/ml; Chronolog Corp, Haverton, PA). The same methods, agonists, and agonist concentrations were used in the OOA cohort with the exception that only one concentration of epinephrine (10 μM) was used and an extra concentration of ADP (5 μM) was tested. FHS tested aggregation for periods of 4 min after administration of ADP (1.0, 3.0, 5.0, and 10.0 μM) and 5 min after administration of epinephrine (0.5, 1.0, 3.0, 5.0 and

10.0 μM); and, lag time to aggregation was assessed after stimulation with 190 μg/ml calf skin-derived type I collagen (Bio/Data Corporation, Horsham, PA). Threshold concentrations to ADP and epinephrine (EC₅₀) were determined as the minimal concentration of agonist required to produce >50% aggregation.

Using an adapted two stage procedure³⁶, platelet aggregation traits were adjusted for age, sex and aspirin-use using linear model, and the residuals from the linear model were inverse normal transformed within each cohort. Given the difference in agonist concentrations used between GeneSTAR/OOA and FHS cohorts, predefined phenotypes were identified and harmonized across studies. For ADP, epinephrine, and collagen independently, identical or closely matching agonist concentrations, the transformed residuals were combined across studies for analysis to test for association between genetic variants and low as well as high concentrations of each agonists. In total, 19 traits were defined: three low-dose ADP traits, four high-dose ADP traits, five low-dose epinephrine traits, three high-dose epinephrine traits, two low-dose collagen traits, and two high-dose collagen traits. Additional details regarding platelet phenotype harmonization are shown in Supplementary Table 1.

TOPMed whole-genome sequencing. WGS was performed to an average depth of 38X using DNA isolated from blood, PCR-free library construction, and Illumina HiSeq X technology. All samples used in this set of TOPMed genomes were from Freeze 5b. Details for variant calling and quality control are described in a companion paper by Taliun et al.³⁷. Briefly, variant discovery and genotype calling was performed jointly, across all the available TOPMed Freeze 5b studies, using the GotCloud pipeline resulting in a single, multi-study, genotype call set. Sample-level quality control was performed to check for pedigree errors, discrepancies between self-reported and genetic sex, and concordance with prior genotyping array data.

Variant annotation. Variant annotation was performed using the WGSa³⁸ and dbNSFP³⁹. Variants were annotated as exonic, splicing, ncRNA, UTR5, UTR3, intronic, upstream, downstream, or intergenic. Exonic variants were further annotated as frameshift insertion, frameshift deletion, frameshift block substitution, stopgain, stoploss, nonframeshift insertion, nonframeshift deletion, nonframeshift block substitution, nonsynonymous variant, synonymous variant, or unknown. Additional scores available included REVEL⁴⁰, MCAP⁴¹ or CADD⁴² effect prediction algorithms.

Single variant tests for association. All analyses in this study were performed on the Analysis Commons⁴³. Variants with minor allele count (MAC) of at least 5 and depth of coverage (DP) of at least 10 were selected for single variant analyses. The GWAS were conducted using GENetic ESTimation and Inference in Structured samples (GENESIS)^{44,45} apps on Analysis Commons. GENESIS uses a linear mixed model with a genetic relationship matrix (GRM) that is robust to population structure and can account for known or cryptic relatedness. The combined transformed residuals were used to conduct null model analysis adjusting for cohort indicators using *genesis_nullmodel* app (<https://github.com/AnalysisCommons/>). Single variant analysis and Sequence Kernel Association Test (SKAT) gene-based analyses were performed using *genesis_tests* app.

We used $p < 5 \times 10^{-8}$ as our genome-wide significant threshold in single variant analysis including conditional analysis for identifying independent signals. Conditional analysis was conducted by selecting the genome-wide significant variant with lowest *p* value on a chromosome for conditioning and performing single variant analysis on the same chromosome. The procedure was repeated until no genome-wide significant variant is identified in conditional analysis by chromosome. Any variant surpassing genome-wide significance in conditional analysis was considered to be an additional signal independent of conditioned variant(s).

Gene-based coding variant tests for association. To improve the power to identify rare variants in coding regions, we aggregated deleterious rare coding variants in 17,774 protein-coding genes and then tested for association with platelet aggregation phenotypes. To enrich for functional variants, only variants with a “deleterious” consequence for its corresponding gene or genes (http://www.ensembl.org/info/genome/variation/predicted_data.html#consequences), were included. For each

protein-coding gene, a set of rare coding variants (MAF < 0.05) was constructed, which was composed of all stop-gain, stop-loss and frameshift variants as well as exonic missense variants that fulfilled one of these criteria: 1) REVEL score > 0.5, 2) M_CAP score was "Deleterious", or 3) CADD score > 30. The protein coding variant groupings were tested using SKAT with the beta-distribution parameters of 1 and 25 as proposed by Wu et al.¹⁹. Significance was evaluated for each platelet aggregation trait after Bonferroni correction ($0.05/17,744 = 2.82 \times 10^{-6}$).

Next we sought to determine which rare deleterious variants in each significant gene were driving the association signal. We iterated through the variants and removed one variant at a time (*leave-one-out* approach) and repeated the SKAT analysis. If a variant made a large contribution to the original association signal, one would expect the signal to significantly weaken with removal of the variant from the gene set.

Super-enhancer based rare variant tests for association. We investigated rare non-coding variants with putative regulatory potential by focusing on megakaryocyte-specific super enhancers (MK SEs). The published MK SEs²⁰ were called based on regions identified as enhancers through genome segmentation across a set of six histone modifications (H3K4me1, H3K4me3, H3K9me3, H3K27ac, H3K27me3 and H3K36me3) profiled in the BLUEPRINT project⁴⁶, aggregating together elements within 12.5 kb and then ranking upon H3K27ac signal with the ROSE algorithm^{47,48}. We annotated rare (MAF < 0.05) non-coding variants located within megakaryocyte DNase I Hypersensitivity Site (DHS) peaks generated by BLUEPRINT and subset to those overlapping with MK SEs. We then applied SKAT, aggregating these non-coding variants on the set of MK SEs ($n = 1065$) to identify the association of these regulatory elements with platelet aggregation phenotypes. Significance was evaluated for each platelet aggregation trait after Bonferroni correction ($0.05/1065 = 4.69 \times 10^{-5}$). When a gene was identified, we conducted *leave-one-out* analysis to identify if a variant(s) contributed to the observed signal.

Replication. Additional samples from each cohort which were not included in TOPMed and therefore not included in the discovery analyses were used to replicate the genome-wide significant variants identified in the discovery analyses. In brief, genotype imputation and replication analyses were conducted by each cohort, and then meta-analysis was used to combine cohort replication analysis results. For signals identified in our gene-based tests, instead of conducting gene-based replication analysis, we replicated the single rare variants that drove the signals and that were identified from *leave-one-out* analyses, as not all selected rare variants had good imputation quality and were available in each cohort⁴⁹. Each cohort independently, and separately by race for GeneSTAR, imputed the 22 autosomes using the TOPMed Freeze5b reference panel with Minimac4⁵⁰. We implemented sample quality control procedures (excluding duplicate/reference samples and gender mismatches) and genotyping quality control procedures (excluding variants with call rate < 95%, HWE p value < 10^{-6} , or MAF < 0.5%). After lifting over to build 38, non-ambiguous strand flips were resolved and ambiguous strand flips were removed. Post-imputation quality control was performed considering 6 MAF bins and using an imputation R^2 cutoff between 0.3 and 0.8, incrementing by 0.1 such that the mean R^2 exceeded 0.8 for each MAF bin. A maximum of 2009 samples were available for replication analyses: 395 OOA, 1289 FHS, 246 GS-EA, and 100 GS-AA.

GS-EA, GS-AA and OOA used the same statistical methods as the discovery analysis for the replication analysis, again using the Analysis Commons. In brief, the platelet traits were adjusted for age, sex and aspirin-use using a linear model separately in each cohort (OOA, GS-EA, GS-AA), and the residuals from these models were then inverse normal transformed within each cohort. Using GENESIS, null models were fitted within each cohort using a linear mixed model with a GRM and no covariates. Using these null models, single variant analyses were performed using GENESIS, again within each cohort. FHS used a linear mixed effects model with a relationship coefficient matrix that accounts for familial correlation implemented in the *coxme*⁵¹ R package to conduct replication analysis, where linear model adjusting for age, sex and aspirin use was used to obtain residuals, the residuals were inverse normalized and then used for testing genetic association. Replication meta-analyses were performed using the sample size weighted approach implemented in METAL⁵². An imputation quality filter of $R^2 \geq 0.7$ was applied prior to meta-analysis, that is, at a particular variant, any cohorts with $R^2 < 0.7$ did not contribute to the variant's meta-analysis. Replication p -values were reported based on a one-sided test since the same effect direction was the expected result to reject the null hypothesis.

Extension of results to an independent GWAS cohort. The Caerphilly Prospective Study (CaPS) participants were relatively healthy, middle aged males at recruitment and their ages at time of blood draw (Exam 2) ranged from 47 to 66. The extent of platelet aggregation to three agonists was measured in PRP adjusted to 300,000 platelets/ μ l with autologous PPP⁵³. Agonists included collagen (42.7 μ g/ml), ADP (0.725 μ M/l), and full-length thrombin (0.056 units/ml). The maximal optical density increase due to platelet aggregation was measured and expressed as a proportion of the difference in optical density between PRP and PPP. Genotyping was performed with the Affymetrix UK Biobank array using the Affymetrix Axiom Analysis program. Following sample and genotyping quality control, imputation

on 22 autosomes was performed using the HRC 1.1 reference panel, resulting in ~7.6 million variants with MAF > 0.01 and $R^2 > 0.4$. GWAS was performed with the Efficient Mixed-Model Association eXpedited (EMMAX) package. For each trait, a linear model was constructed adjusting for age and medication use (anticoagulant, antiplatelet, antilipid, hypoglycemics) and single variant analyses were performed on transformed platelet reactivity values. Maximum sample sizes and phenotype transformations were as follows: ADP ($n = 1177$, natural log), thrombin ($n = 1183$, square root), and collagen ($n = 811$, cube root). Although the agonists differed in some cases in dose or type from the discovery efforts, we had the prior hypothesis that platelet-reactivity increasing alleles for one agonist are likely to be reactivity increasing for other agonists/doses. Note that in CaPS collagen maximal aggregation was measured, and collagen lag time unavailable, thus, the expected effect direction would be opposite to our discovery analyses (as observed for *PEAR1* rs12041331 in Supplementary Table 4 versus collagen lag time discovery results in Supplementary Data 1). Association extension results in CaPS are reported with beta, standard-error, and one-sided p values relative to the hypothesized direction.

Co-localization of expression quantitative trait loci (eQTL) signatures from platelets in GeneSTAR European Americans. A subset of 180 TOPMed GeneSTAR European Americans samples also had RNA-seq data generated using platelets. eQTL analysis was performed as previously described⁵⁴. Here, formal Bayesian co-localization was performed using the *coloc*⁵⁵⁻⁵⁷ package in R for each of the 16 independent loci (Table 1) against all gene transcripts where there was at least one SNP with an eQTL p value $p < 0.003125$ (0.05/16) for the specific gene within 20 KB of the peak variant. This yielded 10 locus-gene pairs (Supplementary Table 4). *coloc* tests five mutually exclusive hypotheses: H0, no GWAS and no eQTL association; H1, association with GWAS, but no eQTL; H2, association with eQTL, but no GWAS; H3, eQTL and GWAS association, but with two independent causal variants; and H4, shared causal variants for both eQTL and GWAS. The main interest is to assess whether there is a shared causal variant between eQTL and GWAS (i.e., H4). The package provides five posterior probabilities for these hypotheses (PP0, PP1, PP2, PP3, and PP4) and PP4 of >75% is considered evidence of a colocalization of GWAS and eQTL. Posterior probabilities for individual variants were evaluated once PP4 was met.

Allele-specific and transcription-factor enriched enhancer assays

Cell culture. K562 is a lymphoblastoid human erythroleukemia cell line derived from a female donor. It is a suspension cell line. K562 cells were cultured and maintained in RPMI 1640 media supplemented with 10% FBS (Sigma-Aldrich), Pen/Strep and L-Gln. Cultures were maintained in a humidified environment at 37 °C with 5% CO₂. K562 cells were passaged every 24–48 h. HEK293 cells were cultured and maintained at low passage in DMEM media supplemented with 5% FBS (Sigma-Aldrich). Cultures were maintained in a humidified environment at 37 °C with 5% CO₂.

Lentivirus production. For Lentivirus production following vectors were used: pInducer-21 lentiviral vector (Addgene), pMD2.G envelope plasmid (Addgene), psPAX2 packaging plasmid (Addgene). 293T-17 cells (ATCC) were cultured and maintained at low passage in DMEM media supplemented with 5% FBS (Sigma-Aldrich). Cultures were maintained in a humidified environment at 37 °C with 5% CO₂. 293T-17 cells were passaged every 24–48 h. Lentiviral plasmids possessing open reading frames of GFP (Empty), POLR2A, NRF1, CTCF, FOSL1, GATA1, GATA2, CEBPB, and NFE2 were cloned into pInducer-21 lentiviral vector. For lentivirus production, 293T-17 cells (ATCC) were transfected with third generation packaging plasmids pMD2.G and psPAX2 (Addgene) and lentiviral plasmids POLR2A, NRF1, CTCF, FOSL1, GATA1, GATA2, CEBPB, and NFE2. Viruses were harvested 48 h post transfections and concentrated by ultracentrifugation at 71,286 \times g for 2 h at 4 °C. Viruses were titrated by serial dilution on 293 T cells using GFP as an indicator.

RNA extraction, reverse transcription, and RT-qPCR. RNA extraction from variously transfected HEK293 and K562 cells was performed using an RNAeasy kit (Qiagen). Reverse transcription was performed using Superscript III (Invitrogen), using Oligo (dT) 15 primer. Quantitative PCRs were performed in triplicate with Taqman primer probe assays, shown in Methods Table 1 and CFX96 real-time PCR detection system (Bio-rad). Target transcript abundance was calculated relative to ACTB (reference gene) using the 2- $\Delta\Delta$ CT method. Gene specific primer pairs are present in methods section Oligonucleotides.

Enhancer function reporter assay. The following vectors were used in this protocol: pGL3 luciferase reporter (Promega), pGL4.74[hRluc/TK] control vector (Promega). ~200–300 base pair non-coding regions of RGS18, ADRA2A and PEAR1 and the associated alleles surrounding the various SNP variants were cloned into the pGL3 luciferase vector. We created two modified constructs to assess functionality of the various loci: wild-type loci carrying no SNP and knock-in of the various SNPs into the respective loci. The constructs were generated via Vectorbuilder. Constructs were sequenced to confirm the expected genotype and to ensure no off-target mutations were introduced. Dual luciferase reporter assays were performed as

described previously with minor modifications^{11,58}. Briefly, GFP, POLR2A, NRF1, CTCF, FOSL1, GATA1, GATA2, CEBPB, or NFE2 overexpressing HEK293 cells or K562 cells were co-transfected with one of the two pGL3 luciferase vectors described above as well pGL3 control according to the manufacturer's instructions. Firefly and Renilla luciferase reporter activity of cell extracts were measured using the Dual-Glo Luciferase Assay System (Promega) on a microplate reader according to the manufacturer's instructions. Each treatment was performed in duplicate and the experiment was repeated three times. Assay primer information is given in Supplementary Table 12.

Phenome-wide association study (PheWAS). The 16 genome-wide significant variants identified from WGS in our discovery cohort (Table 1) as well as significant variants identified in gene-based tests, were examined against clinical phenotypes in the UKBB and BioVU cohorts where available. We queried UKBB GWAS results using SAIGE calculated summary statistics (<http://www.nealelab.is/uk-biobank/faq>). The BioVU repository contains >250,000 DNA samples obtained from discarded blood samples of consented patients at Vanderbilt University Medical Center (Nashville, TN). De-identified DNA samples in the BioVU repository are linked to 1543 clinical diagnostic codes. Of these 1543 clinical diagnostic codes, we identified 71 diagnoses for which platelet function could be in the pathophysiologic pathway to disease expression. Numerous overlapping disease processes were represented among these 71 codes, which we further categorized into the following 6 phenotypes: arterial thrombosis (30 codes), venous thrombosis (3 codes), hypercoagulable state (2 codes), platelet (5 codes), bleeding (26 codes), and anti-thrombotic medication usage (5 codes). The phecodes were matched between UKBB and BioVU for corresponding allelic results.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

TOPMed WGS variant calls are available for all samples through the Database of Genotypes and Phenotypes (dbGaP) under accession number phs001218 for GeneSTAR, phs000956 and phs000391 for OOA and phs000974 for Framingham. Phenotype data for GeneSTAR, OOA and Framingham are also available through this mechanism. Summary statistics are being deposited in the TOPMed GSR (Genomic Summary Results) site. eQTL analysis results used in the co-localization analysis are hosted on a website at: <http://www.biostat.jhsph.edu/~kkammers/GeneSTAR/>.

Received: 8 October 2020; Accepted: 13 April 2021;

Published online: 15 June 2021

References

- Hennekens, C. H., Dyken, M. L. & Fuster, V. Aspirin as a therapeutic agent in cardiovascular diseases: a statement for healthcare professionals from the American Heart Association. *Circulation* **96**, 2751–2753 (1997).
- Jin, R. C., Voetsch, B. & Loscalzo, J. Endogenous mechanisms of inhibition of platelet function. *Microcirculation* **12**, 247–258 (2005).
- Faraday, N. et al. Heritability of platelet responsiveness to aspirin in activation pathways directly and indirectly related to cyclooxygenase-1. *Circulation* **115**, 2490–2496 (2007).
- Johnson, A. D. The genetics of common variation affecting platelet development, function and pharmaceutical targeting. *J. Thromb. Haemost.* **9**, 246–257 (2011).
- O'Donnell, C. J. et al. Genetic and environmental contributions to platelet aggregation: the Framingham heart study. *Circulation* **103**, 3051–3056 (2001).
- Puurunen, M. K. et al. ADP platelet hyperreactivity predicts cardiovascular disease in the FHS (Framingham Heart Study). *J Am Heart Assoc* **7**, <https://doi.org/10.1161/JAHA.118.008522> (2018).
- Qayyum, R. et al. Greater collagen-induced platelet aggregation following cyclooxygenase 1 inhibition predicts incident acute coronary syndromes. *Clin. Transl. Sci.* **8**, 17–22 (2015).
- Chen, M. H. et al. Exome-chip meta-analysis identifies association between variation in ANKRD26 and platelet aggregation. *Platelets* **30**, 164–173 (2019).
- Eicher, J. D. et al. Whole exome sequencing in the Framingham Heart Study identifies rare variation in HYAL2 that influences platelet aggregation. *Thromb. Haemost.* **117**, 1083–1092 (2017).
- Johnson, A. D. et al. Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. *Nat. Genet* **42**, 608–613 (2010).
- Rodriguez, B. A. T. et al. A platelet function modulator of thrombin activation is causally linked to cardiovascular disease and affects PAR4 receptor signaling. *Am. J. Hum. Genet* **107**, 211–221 (2020).
- Keramati, A. R. et al. Targeted deep sequencing of the PEAR1 locus for platelet aggregation in European and African American families. *Platelets* **30**, 380–386 (2019).
- Eicher, J. D., Xue, L., Ben-Shlomo, Y., Beswick, A. D. & Johnson, A. D. Replication and hematological characterization of human platelet reactivity genetic associations in men from the Caerphilly Prospective Study (CaPS). *J. Thromb. Thrombolysis* **41**, 343–350 (2016).
- Faraday, N. et al. Identification of a specific intronic PEAR1 gene variant associated with greater platelet aggregability and protein expression. *Blood* **118**, 3367–3375 (2011).
- Herrera-Galeano, J. E. et al. A novel variant in the platelet endothelial aggregation receptor-1 gene is associated with increased platelet aggregability. *Arterioscler Thromb. Vasc. Biol.* **28**, 1484–1490 (2008).
- Delesque-Touchard, N. et al. Regulator of G-protein signaling 18 controls both platelet generation and function. *PLoS One* **9**, e113215 (2014).
- Alshool, F. Z. et al. The regulator of G-protein signaling 18 regulates platelet aggregation, hemostasis and thrombosis. *Biochem Biophys. Res. Commun.* **462**, 378–382 (2015).
- Ma, P. et al. Modulating platelet reactivity through control of RGS18 availability. *Blood* **126**, 2611–2620 (2015).
- Wu, M. C. et al. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet* **89**, 82–93 (2011).
- Petersen, R. et al. Platelet function is modified by common sequence variation in megakaryocyte super enhancers. *Nat. Commun.* **8**, 16058 (2017).
- Kim, Y. et al. Targeted deep resequencing identifies coding variants in the PEAR1 gene that play a role in platelet aggregation. *PLoS One* **8**, e64179 (2013).
- Izzi, B. et al. Allele-specific DNA methylation reinforces PEAR1 enhancer activity. *Blood* **128**, 1003–1012 (2016).
- Nanda, N. et al. Platelet endothelial aggregation receptor 1 (PEAR1), a novel epidermal growth factor repeat-containing transmembrane receptor, participates in platelet contact-induced activation. *J. Biol. Chem.* **280**, 24680–24689 (2005).
- Izzi, B., Noro, F., Cludts, K., Freson, K. & Hoylaerts, M. F. Cell-specific PEAR1 methylation studies reveal a locus that coordinates expression of multiple genes. *Int J Mol Sci* **19**, <https://doi.org/10.3390/ijms19041069> (2018).
- Samuelov, L. et al. SVEP1 plays a crucial role in epidermal differentiation. *Exp. Dermatol* **26**, 423–430 (2017).
- Karpanen, T. et al. An evolutionarily conserved role for polydom/Svep1 during lymphatic vessel formation. *Circ. Res.* **120**, 1263–1275 (2017).
- Morooka, N. et al. Polydom is an extracellular matrix protein involved in lymphatic vessel remodeling. *Circ. Res.* **120**, 1276–1288 (2017).
- Eicher, J. D. et al. Characterization of the platelet transcriptome by RNA sequencing in patients with acute myocardial infarction. *Platelets* **27**, 230–239 (2016).
- Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
- Sato-Nishiuchi, R. et al. Polydom/SVEP1 is a ligand for integrin alpha9beta1. *J. Biol. Chem.* **287**, 25615–25630 (2012).
- Winkler, M. J. et al. Functional investigation of the coronary artery disease gene SVEP1. *Basic Res. Cardiol.* **115**, 67 (2020).
- Stuckey, T. D. et al. Impact of aspirin and clopidogrel hypo-responsiveness in patients treated with drug-eluting stents: 2-year results of a prospective, multicenter registry study. *JACC Cardiovasc. Inter.* **10**, 1607–1617 (2017).
- Price, M. J. et al. Platelet reactivity and cardiovascular outcomes after percutaneous coronary intervention: a time-dependent analysis of the Gauging Responsiveness with a VerifyNow P2Y12 assay: Impact on Thrombosis and Safety (GRAVITAS) trial. *Circulation* **124**, 1132–1137 (2011).
- Bray, P. F. et al. Heritability of platelet function in families with premature coronary artery disease. *J. Thromb. Haemost.* **5**, 1617–1623 (2007).
- Shuldiner, A. R. et al. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA* **302**, 849–857 (2009).
- Sofer, T. et al. A fully adjusted two-stage procedure for rank-normalization in genetic association studies. *Genet Epidemiol.* **43**, 263–275 (2019).
- Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
- Liu, X. et al. WGS: an annotation pipeline for human genome sequencing studies. *J. Med Genet* **53**, 111–112 (2016).
- Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum. Mutat.* **37**, 235–241 (2016).
- Ioannidis, N. M. et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet* **99**, 877–885 (2016).
- Jagadeesh, K. A. et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet* **48**, 1581–1586 (2016).

42. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet* **46**, 310–315 (2014).
43. Brody, J. A. et al. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nat. Genet* **49**, 1560–1563 (2017).
44. Conomos, M. P., Miller, M. B. & Thornton, T. A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol.* **39**, 276–293 (2015).
45. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
46. Stunnenberg, H. G. International Human Epigenome Consortium; Martin Hirst. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell* **167**, 1145–1149 (2016).
47. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
48. Loven, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
49. Van Hout, C. V. et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).
50. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet* **48**, 1284–1287 (2016).
51. Ripatti, S. & Palmgren, J. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*. **56**, 1016–1022 (2000).
52. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
53. Elwood, P. C. et al. Ischemic heart disease and platelet aggregation. The Caerphilly Collaborative Heart Disease Study. *Circulation* **83**, 38–44 (1991).
54. Kammers, K. et al. Transcriptional profile of platelets and iPSC-derived megakaryocytes from whole genome and RNA sequencing. *Blood*, <https://doi.org/10.1182/blood.2020006115> (2020).
55. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
56. Wallace, C. Statistical testing of shared genetic control for potentially related traits. *Genet Epidemiol.* **37**, 802–813 (2013).
57. Wallace, C. et al. Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum. Mol. Genet* **21**, 2815–2824 (2012).
58. Bhan, A. et al. Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol. *J. Mol. Biol.* **425**, 3707–3722 (2013).
59. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).

Acknowledgements

Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for GeneSTAR (Genetic Study of Atherosclerosis Risk) was performed at Macrogen, Illumina, and the Broad Institute of MIT and Harvard (HHSN268201500014C). WGS for the Old Order Amish (Genetics of Cardiometabolic Health in the Amish) was performed at the Broad Institute of MIT and Harvard (3R01HL121007-01S1). WGS for The Framingham Heart Study (Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1). For the Old Order Amish this investigation was supported by National Institutes of Health grants U01 GM074518, U01 HL105198, R01 HL137922, R01 HL121007, and the University of Maryland Mid-Atlantic Nutrition and Obesity Research Center (P30 DK072488). GeneSTAR was supported by the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518, HL087698, HL112064, HL11006, HL118356) and by a grant from the National Institutes of Health/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center. The Framingham Heart Study is conducted and supported by the NHLBI in collaboration with Boston University (Contract No. N01-HC-25195), its contract with Affymetrix, Inc., for genome-wide genotyping services (Contract No. N02-HL-6-4278 and Contract No. HHSN268201500001I). MHC, BATC and ADJ were supported by NHLBI Intramural funding. The Caerphilly Prospective study was undertaken by the former MRC Epidemiology Unit (South Wales) and was funded by the Medical Research Council of the UK. The data archive is maintained by the School of Social and Community Medicine, University of Bristol. This study makes use of data generated by the

BLUEPRINT Consortium. A full list of the investigators who contributed to the generation of the data is available from www.blueprint-epigenome.eu. Funding for the project was provided by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 282510 BLUEPRINT. Additional support came from the National Blood Foundation/American Association of Blood Banks (FP01021164), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; U54DK110805) and the National Research Service Award (NRSA)'s Joint Program in Transfusion Medicine (T32 4T32HL066987-15 to A.B.). BioVU resource analyses were supported by National Institutes of Health/National Genome Research Institute grant U01HG009086. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Author contributions

A.R.K., M.-H.C., B.A.T.R. led equal roles in analysis, interpretation and writing of the paper. J.P.L., R.A.M., A.D.J. led equal senior roles for the study. A.R.K. wrote the first draft of manuscript with contribution and editing from M.-H.C., B.A.T.R., L.R.Y., M.A.T., J.A.B., L.C.B., N.F., J.P.L., R.A.M. and A.D.J. Genome wide analyses were performed by A.R.K., B.A.T.R., B.J.G., L.R.Y., M.-H.C. and K.R. RNA-sequencing and eQTL analyses were performed by R.A.M., Kai.K., M.A.T., I.R., L.R.Y., A.R.K., Kan.K. and K.I. Imputation of genomic data and replication analyses were performed by M.-H.C., B.A.T.R., L.R.Y., B.J.G., A.P., L.A.C., and M.H.K. L.R.Y., N.F., L.C.B. and R.A.M. were involved in the guidance, collection and analysis for Genetic Study of Atherosclerosis Risk (GeneSTAR) phenotype data. B.J.G., K.R., B.D.M., J.P.L., J.R.O. and A.R.S. were involved in the guidance, collection and analysis for Older Order Amish Study (OAA) phenotype data. B.A.T.R., M.-H.C. and A.D.J. were involved in the guidance and analysis for Framingham Heart Study (FHS) phenotype data. T.M.S. and A.B. established the imMKCL system, and A.B. designed all functional experiments with input from B.A.T.R. and A.D.J. and carried them out. X.Z., Q.W. and B.L. carried out BioVU pheWAS analyses. A.D.J. funded genotyping of the CaPS cohort. M.-H.C. performed genotype QC, calling and imputation of the CaPS cohort with input from A.D.J., B.A.T.R., M.-H.C. and A.D.J. conducted CaPS genotype-phenotype analyses.

Funding

Open Access funding provided by the National Institutes of Health (NIH).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23470-9>.

Correspondence and requests for materials should be addressed to J.P.L., R.A.M. or A.D.J.

Peer review information *Nature Communications* thanks Willem Ouwehand and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2021

NHLBI Trans-Omics for Precision (TOPMed) Consortium

Namiko Abe¹⁹, Goncalo Abecasis²⁰, Francois Aguet²¹, Christine Albert²², Laura Almasy²³, Alvaro Alonso²⁴, Seth Ament²⁵, Peter Anderson²⁶, Pramod Anugu²⁷, Deborah Applebaum-Bowden²⁸, Kristin Ardlie²¹, Dan Arking²⁹, Donna K. Arnett³⁰, Allison Ashley-Koch³¹, Stella Aslibekyan³², Tim Assimes³³, Paul Auer³⁴, Dimitrios Avramopoulos²⁹, Najib Ayas³⁵, Adithya Balasubramanian³⁶, John Barnard³⁷, Kathleen Barnes³⁸, R. Graham Barr³⁹, Emily Barron-Casella²⁹, Lucas Barwick⁴⁰, Terri Beaty²⁹, Gerald Beck³⁷, Diane Becker²⁹, Lewis Becker²⁹, Rebecca Beer⁴¹, Amber Beitelshes²⁵, Emelia Benjamin⁴², Takis Benos⁴³, Marcos Bezerra⁴⁴, Larry Bielak²⁰, Joshua Bis²⁶, Thomas Blackwell²⁰, John Blangero⁴⁵, Eric Boerwinkle⁴⁶, Donald W. Bowden⁴⁷, Russell Bowler⁴⁸, Jennifer Brody²⁶, Ulrich Broeckel⁴⁹, Jai Broome²⁶, Deborah Brown⁴⁶, Karen Bunting¹⁹, Esteban Burchard⁵⁰, Carlos Bustamante³³, Erin Buth²⁶, Brian Cade⁵¹, Jonathan Cardwell⁵², Vincent Carey⁵¹, Julie Carrier⁵³, Cara Carty⁵⁴, Richard Casaburi⁵⁵, Juan P. Casas Romero⁵¹, James Casella²⁹, Peter Castaldi⁵¹, Mark Chaffin²¹, Christy Chang²⁵, Yi-Cheng Chang⁵⁶, Daniel Chasman⁵¹, Sameer Chavan⁵², Bo-Juen Chen¹⁹, Wei-Min Chen⁵⁷, Yii-Derlda Chen⁵⁸, Michael Cho⁵¹, Seung Hoan Choi²¹, Lee-Ming Chuang⁵⁶, Mina Chung³⁷, Ren-Hua Chung⁵⁹, Clary Clish²¹, Suzy Comhair³⁷, Matthew Conomos²⁶, Elaine Cornell⁶⁰, Adolfo Correa²⁷, Carolyn Crandall⁵⁵, James Crapo⁴⁸, L. Adrienne Cupples⁶¹, Joanne Curran⁴⁵, Jeffrey Curtis²⁰, Brian Custer⁶², Coleen Damcott²⁵, Dawood Darbar⁶³, Sean David⁶⁴, Colleen Davis²⁶, Michelle Daya⁵², Mariza de Andrade⁶⁵, Lisa de las Fuentes⁶⁶, Paul de Vries⁴⁶, Michael DeBaun⁶⁷, Ranjan Deka⁶⁸, Dawn DeMeo⁵¹, Scott Devine²⁵, Huyen Dinh³⁶, Harsha Doddapaneni³⁶, Qing Duan⁶⁹, Shannon Dugan-Perez³⁶, Ravi Duggirala⁷⁰, Jon Peter Durda⁶⁰, Susan K. Dutcher⁶⁶, Charles Eaton⁷¹, Lynette Ekunwe²⁷, Adel El Boueiz⁷², Patrick Ellinor⁷³, Leslie Emery²⁶, Serpil Erzurum³⁷, Charles Farber⁵⁷, Jesse Farek³⁶, Tasha Fingerlin⁴⁸, Matthew Flickinger²⁰, Myriam Fornage⁴⁶, Nora Franceschini⁶⁹, Chris Frazar²⁶, Mao Fu²⁵, Stephanie M. Fullerton²⁶, Lucinda Fulton⁶⁶, Stacey Gabriel²¹, Weiniu Gan⁴¹, Shanshan Gao⁵², Yan Gao²⁷, Margery Gass⁷⁴, Heather Geiger¹⁹, Bruce Gelb⁷⁵, Mark Geraci⁴³, Soren Germer¹⁹, Robert Gerszten⁷⁶, Auyon Ghosh⁵¹, Richard Gibbs³⁶, Chris Gignoux³³, Mark Gladwin⁴³, David Glahn⁷⁷, Stephanie Gogarten²⁶, Da-Wei Gong²⁵, Harald Goring⁷⁸, Sharon Graw³⁸, Kathryn J. Gray⁷⁹, Daniel Grine⁵², Colin Gross²⁰, C.Charles Gu⁶⁶, Yue Guan²⁵, Xiuqing Guo⁵⁸, Namrata Gupta²¹, David M. Haas⁸⁰, Jeff Haessler⁷⁵, Michael Hall²⁷, Yi Han³⁶, Patrick Hanly⁸¹, Daniel Harris⁸², Nicola L. Hawley⁸³, Jiang He⁸⁴, Ben Heavner²⁶, Susan Heckbert²⁶, Ryan Hernandez⁵⁰, David Herrington⁴⁷, Craig Hersh⁵¹, Bertha Hidalgo³², James Hixson⁴⁶, Brian Hobbs⁵¹, John Hokanson⁵², Elliott Hong²⁵, Karin Hoth⁸⁵, Chao Agnes Hsiung⁵⁹, Jianhong Hu³⁶, Yi-Jen Hung⁸⁶, Haley Huston⁸⁷, Chii Min Hwu⁸⁸, Marguerite Ryan Irvin³², Rebecca Jackson⁸⁹, Deepti Jain²⁶, Cashell Jaquish⁴¹, Jill Johnsen⁸⁷, Andrew Johnson⁴¹, Craig Johnson²⁶, Rich Johnston²⁴, Kimberly Jones²⁹, Hyun Min Kang²⁰, Robert Kaplan⁹⁰, Sharon Kardia²⁰, Shannon Kelly⁵⁰, Eimear Kenny⁷⁵, Michael Kessler²⁵, Alyna Khan²⁶, Ziad Khan³⁶, Wonji Kim⁷², John Kimoff⁹¹, Greg Kinney⁹³, Barbara Konkle⁸⁷, Charles Kooperberg⁷⁵, Holly Kramer⁹⁴, Christoph Lange⁹², Ethan Lange⁵², Leslie Lange⁵², Cathy Laurie²⁶, Cecelia Laurie²⁶, Meryl LeBoff⁵¹, Jiwon Lee⁵¹, Sandra Lee³⁶, Wen-Jane Lee⁸⁸, Jonathon LeFaive²⁰, David Levine²⁶, Dan Levy⁴¹, Joshua Lewis²⁵, Xiaohui Li⁸⁸, Yun Li⁶⁹, Henry Lin⁵⁸, Honghuang Lin⁶¹, Xihong Lin⁹², Simin Liu⁷¹, Yongmei Liu³¹, Yu Liu³³, Ruth J. F. Loos⁷⁵, Steven Lubitz⁷², Kathryn Lunetta⁶¹, James Luo⁴¹, Ulysses Magalang⁹⁵, Michael Mahaney⁴⁵, Barry Make²⁹, Ani Manichaikul⁵⁸, Alisa Manning⁹⁶, JoAnn Manson⁵¹, Lisa Martin⁹⁷, Melissa Marton¹⁹, Susan Mathai⁵², Rasika Mathias²⁹, Susanne May²⁶, Patrick McArdle²⁵, Merry-Lynn McDonald³², Sean McFarland⁷², Stephen McGarvey⁷¹, Daniel McGoldrick²⁶, Caitlin McHugh²⁶, Becky McNeil⁹⁸, Hao Mei²⁷, James Meigs⁷², Vipin Menon³⁶, Luisa Mestroni³⁸, Ginger Metcalf³⁶, Deborah A. Meyers⁹⁹, Emmanuel Mignot¹⁰⁰, Julie Mikulla⁴¹, Nancy Min²⁷, Mollie Minear¹⁰¹, Ryan L. Minster⁴³, Braxton D. Mitchell²⁵, Matt Moll⁵¹, Zeineen Momin³⁶, May E. Montasser²⁵, Courtney Montgomery¹⁰², Donna Muzny³⁶, Josyf C. Mychaleckyj⁵⁷, Girish Nadkarni⁷⁶, Rakhi Naik²⁹,

Take Naseri¹⁰³, Pradeep Natarajan²¹, Sergei Nekhai¹⁰⁴, Sarah C. Nelson²⁶, Bonnie Neltner⁵², Caitlin Nessner³⁶, Deborah Nickerson²⁶, Osuji Nkechinyere³⁶, Kari North⁶⁹, Jeff O'Connell¹⁰⁵, Tim O'Connor²⁵, Heather Ochs-Balcom¹⁰⁶, Geoffrey Okwuonu³⁶, Allan Pack¹⁰⁷, David T. Paik³³, Nicholette Palmer⁴⁷, James Pankow¹⁰⁸, George Papanicolaou⁴¹, Cora Parker¹⁰⁹, Gina Peloso⁶¹, Juan Manuel Peralta⁶⁹, Marco Perez³³, James Perry²⁵, Ulrike Peters⁷³, Patricia Peyser²⁰, Lawrence S. Phillips²⁴, Jacob Pleiness²⁰, Toni Pollin²⁵, Wendy Post²⁹, Julia Powers Becker⁵², Meher Preethi Boorgula⁵², Michael Preuss⁷⁵, Bruce Psaty²⁶, Pankaj Qasba⁴¹, Dandi Qiao⁵¹, Zhaohui Qin²⁴, Nicholas Rafaels⁵², Laura Raffield⁶⁹, Mahitha Rajendran³⁶, Vasan S. Ramachandran⁶², D. C. Rao⁶⁶, Laura Rasmussen-Torvik¹¹⁰, Aakrosh Ratan⁵⁸, Susan Redline⁵¹, Robert Reed²⁵, Catherine Reeves¹⁹, Elizabeth Regan⁴⁸, Alex Reiner¹¹¹, Muagututi'a Sefuiva Reupena¹¹², Ken Rice²⁶, Stephen Rich⁵⁸, Rebecca Robillard¹¹³, Nicolas Robine¹⁹, Dan Roden⁶⁷, Carolina Roselli²¹, Jerome Rotter⁵⁸, Ingo Ruczinski²⁹, Alexi Runnels¹⁹, Pamela Russell⁵², Sarah Ruuska⁸⁷, Kathleen Ryan²⁵, Ester Cerdeira Sabino¹¹⁴, Danish Saleheen³⁹, Shabnam Salimi²⁵, Sejal Salvi³⁶, Steven Salzberg²⁹, Kevin Sandow⁵⁸, Vijay G. Sankaran⁷², Jireh Santibanez³⁶, Karen Schwander⁶⁶, David Schwartz⁵², Frank Sciruba⁴³, Christine Seidman¹¹⁵, Jonathan Seidman¹¹⁵, Frederic Series¹¹⁶, Vivien Sheehan²⁴, Stephanie L. Sherman²⁴, Amol Shetty²⁵, Aniket Shetty⁵², Wayne Hui-Heng Sheu⁸⁸, M. Benjamin Shoemaker⁶⁷, Brian Silver¹¹⁷, Edwin Silverman⁵¹, Robert Skomro¹¹⁸, Albert Vernon Smith²⁰, Jennifer Smith²⁰, Josh Smith²⁶, Nicholas Smith²⁶, Tanja Smith¹⁹, Sylvia Smoller⁹⁰, Beverly Snively⁴⁷, Michael Snyder³³, Tamar Sofer⁵¹, Nona Sotoodehnia²⁶, Adrienne M. Stimp²⁶, Garrett Storm⁵², Elizabeth Streeten²⁵, Jessica Lasky Su⁵¹, Yun Ju Sung⁶⁷, Jody Sylvia⁵¹, Adam Szpiro²⁶, Daniel Taliun²⁰, Hua Tang³³, Margaret Taub²⁹, Kent D. Taylor⁵⁸, Matthew Taylor³⁸, Simeon Taylor²⁵, Marilyn Telen³¹, Timothy A. Thornton²⁶, Machiko Threlkeld²⁶, Lesley Tinker⁷³, David Tirschwell²⁶, Sarah Tishkoff¹⁰⁷, Hemant Tiwari³², Catherine Tong²⁶, Russell Tracy⁶⁰, Michael Tsai¹⁰⁸, Dhananjay Vaidya²⁹, David Van Den Berg¹¹⁹, Peter VandeHaar²⁰, Scott Vrieze¹⁰⁸, Tarik Walker⁵², Robert Wallace⁸⁵, Avram Walts⁵², Fei Fei Wang²⁶, Heming Wang¹²⁰, Jiongming Wang²⁰, Karol Watson⁵⁵, Jennifer Watt³⁶, Daniel E. Weeks⁴³, Bruce Weir²⁶, Scott T. Weiss⁵¹, Lu-Chen Weng⁷², Jennifer Wessel⁸⁰, Cristen Willer²⁰, Kayleen Williams²⁶, L. Keoki Williams¹²¹, Carla Wilson⁵¹, James Wilson¹²², Lara Winterkorn¹⁹, Quenna Wong²⁶, Joseph Wu³³, Huichun Xu²⁵, Lisa Yanek²⁹, Ivana Yang⁵², Ketian Yu²⁰, Seyedeh Maryam Zekavat²¹, Yingze Zhang⁴³, Snow Xueyan Zhao⁴⁸, Wei Zhao²⁰, Xiaofeng Zhu¹²³, Michael Zody¹⁹ & Sebastian Zoellner²⁰

¹⁹New York Genome Center, New York, NY, USA. ²⁰University of Michigan, Ann Arbor, MI, USA. ²¹Broad Institute, Cambridge, MA, USA. ²²Cedars Sinai, Boston, MA, USA. ²³Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA, USA. ²⁴Emory University, Atlanta, GA, USA. ²⁵University of Maryland, Baltimore, MD, USA. ²⁶University of Washington, Seattle, WA, USA. ²⁷University of Mississippi, Jackson, MS, USA. ²⁸National Institutes of Health, Bethesda, MD, USA. ²⁹Johns Hopkins University, Baltimore, MD, USA. ³⁰University of Kentucky, Lexington, KY, USA. ³¹Duke University, Durham, NC, USA. ³²University of Alabama, Birmingham, LA, USA. ³³Stanford University, Stanford, CA, USA. ³⁴University of Wisconsin Milwaukee, Milwaukee, WI, USA. ³⁵Providence Health Care, Vancouver, BC, Canada. ³⁶Baylor College of Medicine Human Genome Sequencing Center, Houston, TX, USA. ³⁷Cleveland Clinic, Cleveland, OH, USA. ³⁸University of Colorado Anschutz Medical Campus, Aurora, OH, USA. ³⁹Columbia University, New York, NY, USA. ⁴⁰The Emmes Corporation, Rockville, MD, USA. ⁴¹National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA. ⁴²Boston University, Massachusetts General Hospital, Boston, MA, USA. ⁴³University of Pittsburgh, Pittsburgh, PA, USA. ⁴⁴Fundação de Hematologia e Hemoterapia de Pernambuco - Hemope, Recife, BR, Brazil. ⁴⁵University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA. ⁴⁶University of Texas Health at Houston, Houston, TX, USA. ⁴⁷Wake Forest Baptist Health, Winston-Salem, MA, USA. ⁴⁸National Jewish Health, Denver, CO, USA. ⁴⁹Medical College of Wisconsin, Milwaukee, WI, USA. ⁵⁰University of California, San Francisco, San Francisco, CA, USA. ⁵¹Brigham & Women's Hospital, Boston, MA, USA. ⁵²University of Colorado at Denver, Denver, CO, USA. ⁵³University of Montreal, Montreal, QC, Canada. ⁵⁴Washington State University, Pullman, WA, USA. ⁵⁵University of California, Los Angeles, Los Angeles, CA, USA. ⁵⁶National Taiwan University, Taipei, TW, Taiwan. ⁵⁷University of Virginia, Charlottesville, VA, USA. ⁵⁸Lundquist Institute, Torrance, CA, USA. ⁵⁹National Health Research Institute Taiwan, Miaoli County, TW, Taiwan. ⁶⁰University of Vermont, Burlington, VT, USA. ⁶¹Boston University, Boston, MA, USA. ⁶²Vitalant Research Institute, San Francisco, CA, USA. ⁶³University of Illinois at Chicago, Chicago, IL, USA. ⁶⁴University of Chicago, Chicago, IL, USA. ⁶⁵Mayo Clinic, Rochester, NY, USA. ⁶⁶Washington University in St Louis, St. Louis, WA, USA. ⁶⁷Vanderbilt University, Nashville, TN, USA. ⁶⁸University of Cincinnati, Cincinnati, OH, USA. ⁶⁹University of North Carolina, Chapel Hill, NC, USA. ⁷⁰University of Texas Rio Grande Valley School of Medicine, Edinburg, TX, USA. ⁷¹Brown University, Providence, RI, USA. ⁷²Harvard University, Cambridge, MA, USA. ⁷³Massachusetts General Hospital, Boston, MA, USA. ⁷⁴Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ⁷⁵Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁷⁶Beth Israel Deaconess Medical Center, Boston, MA, USA. ⁷⁷Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. ⁷⁸University of Texas Rio Grande Valley School of Medicine, San Antonio,

TX, USA. ⁷⁹Mass General Brigham, Boston, MA, USA. ⁸⁰Indiana University, Indianapolis, IN, USA. ⁸¹University of Calgary, Calgary, AB, Canada. ⁸²University of Maryland, Philadelphia, PA, USA. ⁸³Yale University, New Haven, CT, USA. ⁸⁴Tulane University, New Orleans, LA, USA. ⁸⁵University of Iowa, Iowa City, IA, USA. ⁸⁶Tri-Service General Hospital National Defense Medical Center, Taichung City, TW, Taiwan. ⁸⁷Blood Works Northwest, Seattle, WA, USA. ⁸⁸Taichung Veterans General Hospital Taiwan, Taichung City, TW, Taiwan. ⁸⁹Oklahoma State University Medical Center, Columbus, OK, USA. ⁹⁰Albert Einstein College of Medicine, New York, NY, USA. ⁹¹McGill University, Montréal, Quebec, CA, USA. ⁹²Harvard School of Public Health, Boston, MA, USA. ⁹³University of Colorado at Denver, Aurora, CO, USA. ⁹⁴Loyola University, Maywood, IL, USA. ⁹⁵Ohio State University, Columbus, OH, USA. ⁹⁶Broad Institute, Harvard University, Massachusetts General Hospital, Cambridge, MA, USA. ⁹⁷George Washington University, Washington, WA, USA. ⁹⁸RTI International, Washington, WA, USA. ⁹⁹University of Arizona, Tucson, AZ, USA. ¹⁰⁰Stanford University, Palo Alto, CA, USA. ¹⁰¹National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA. ¹⁰²Oklahoma Medical Research Foundation, Oklahoma City, OK, USA. ¹⁰³Ministry of Health, Government of Samoa, Apia, WS, Samoa. ¹⁰⁴Howard University, Washington, WA, USA. ¹⁰⁵University of Maryland, Baltimore, MD, USA. ¹⁰⁶University at Buffalo, Buffalo, NY, USA. ¹⁰⁷University of Pennsylvania, Philadelphia, PA, USA. ¹⁰⁸University of Minnesota, Minneapolis, MN, USA. ¹⁰⁹RTI International, Research Triangle Park, NC, USA. ¹¹⁰Northwestern University, Chicago, IL, USA. ¹¹¹Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA, USA. ¹¹²Lutia I Puava Ae Mapu I Fagalele, Apia, WS, Samoa. ¹¹³University of Ottawa, Ottawa, QC, Canada. ¹¹⁴Universidade de Sao Paulo, Sao Paulo, BR, Brazil. ¹¹⁵Harvard Medical School, Boston, MA, USA. ¹¹⁶Université Laval, Quebec City, CA, USA. ¹¹⁷UMass Memorial Medical Center, Worcester, MA, USA. ¹¹⁸University of Saskatchewan, Saskatoon, SK, Canada. ¹¹⁹University of Southern California, Los Angeles, CA, USA. ¹²⁰Brigham & Women's Hospital, Mass General Brigham, Boston, MA, USA. ¹²¹Henry Ford Health System, Detroit, MI, USA. ¹²²Beth Israel Deaconess Medical Center, Cambridge, MA, USA. ¹²³Case Western Reserve University, Cleveland, OH, USA.