HHS Public Access

Author manuscript

Thromb Haemost. Author manuscript; available in PMC 2020 September 04.

Published in final edited form as:

Thromb Haemost. 2017 June 02; 117(6): 1083–1092. doi:10.1160/TH16-09-0677.

Whole exome sequencing in the Framingham Heart Study identifies rare variation in *HYAL2* that influences platelet aggregation

John D. Eicher[1,2], **Ming-Huei Chen**[1,2], **Achilleas N. Pitsillides**[2], **Honghuang Lin**[3], **Narayanan Veeraraghavan**[4], **Jennifer A. Brody**[5], **Ginger A. Metcalf**[6], **Donna M. Muzny**[6], **Richard A. Gibbs**[6], **Diane M. Becker**[7], **Lewis C. Becker**[7,8], **Nauder Faraday**[9], **Rasika A. Mathias**[10], **Lisa R. Yanek**[7], **Eric Boerwinkle**[6,7], **L. Adrienne Cupples**[2,8], **Andrew D. Johnson**[1,2]

[1]Population Sciences Branch, National Heart, Lung, Blood Institute, National Institutes of Health, Framingham, MA, USA 01702

[2]The Framingham Heart Study, Framingham, MA, USA 01702

[3]Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA 02118

[4]Rady Children's Institute for Genomic Medicine, San Diego, CA 92123

[5]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA, 98101

[6]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA 77030

[7]GeneSTAR Research Program, Department of Medicine, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD USA 21287

[8]GeneSTAR Research Program, Department of Medicine, Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD USA 21287

[9]GeneSTAR Research Program, Department of Anesthesiology & Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD USA 21287

[10]GeneSTAR Research Program, Department of Medicine, Division of Allergy and Clinical Immunology, Johns Hopkins University School of Medicine, Baltimore, MD USA 21287

Abstract

Corresponding Author: Andrew D. Johnson, Tenure Track Investigator, Population Sciences Branch, National Heart, Lung, and Blood Institute, The Framingham Heart Study, 73 Mt. Wayte Ave. Suite #2, Framingham, MA 01702, johnsonad2@nhlbi.nih.gov, Phone: (508)-663-4082.

[11]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA 77030

[12]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118

Author Contributions

JDE and MHC performed the statistical analyses in Framingham; ANP and HL analyzed, annotated, and managed Framingham sequencing data; NV, JAB, GAM, DMM, and RAG generated and analyzed original sequencing data; EB and LAC conceived CHARGE exome sequencing study; LRY, DMB, LCB, NF, and RAM conceived and executed the GeneSTAR whole genome sequencing study; LRY did statistical analyses of GeneSTAR; JDE and ADJ wrote the manuscript with all authors editing and approving; JDE, MHC, and ADJ designed the overall research study.

Conflict of Interests

The authors declare no competing financial interests.

Inhibition of platelet reactivity is a common therapeutic strategy in secondary prevention of cardiovascular disease. Genetic and environmental factors influence inter-individual variation in platelet reactivity. Identifying genes that contribute to platelet reactivity can reveal new biological mechanisms and possible therapeutic targets. Here, we examined rare coding variation to identify genes associated with platelet reactivity in a population-based cohort. To do so, we performed whole exome sequencing in the Framingham Heart Study and conducted single variant and gene-based association tests against platelet reactivity to collagen, adenosine diphosphate (ADP), and epinephrine agonists in up to 1,211 individuals. Single variant tests revealed no significant associations ($p<1.44\times10^{-7}$), though we observed a suggestive association with previously implicated *MRVI1* (rs11042902, $p=1.95\times10^{-7}$). Using gene-based association tests of rare and low-frequency variants, we found significant associations of *HYAL2* with increased ADP-induced aggregation ($p=1.07\times10^{-7}$) and *GSTZ1* with increased epinephrine-induced aggregation ($p=1.62\times10^{-6}$). *HYAL2* also showed suggestive associations with epinephrine-induced aggregation ($p=2.64\times10^{-5}$). The rare variants in the *HYAL2* gene-based association included a missense variant (N357S) at a known N-glycosylation site and a nonsense variant (Q406*) that removes a glycophosphatidylinositol (GPI) anchor from the resulting protein. These variants suggest that improper membrane trafficking of HYAL2 influences platelet reactivity. We also observed suggestive associations of *AR* ($p=7.39\times10^{-6}$) and *MAPRE1* ($p=7.26\times10^{-6}$) with ADP-induced reactivity. Our study demonstrates that gene-based tests and other grouping strategies of rare variants are powerful approaches to detect associations in population-based analyses of complex traits not detected by single variant tests and possible new genetic influences on platelet reactivity.

Introduction

Platelets are circulating anucleate blood cells derived from megakaryocytes in the bone marrow. The primary role of platelets is to respond to injury in the vasculature by aggregating at the wound site and initiating the blood clotting and repair processes. Platelet reactivity to injury and agonists varies across the population(1–3). More reactive platelets can lead to excessive or inappropriate clotting, while less reactive platelets may lead to impaired thrombosis and repair. Not only does platelet reactivity directly impact clotting, but it also contributes to risk of developing cardiovascular disease and individuals' prognosis(4). Controlling platelet reactivity levels with anti-platelet medications such as aspirin and clopidogrel is a common strategy for secondary prevention of cardiovascular disease(5). Elucidating other factors that modulate platelet reactivity will uncover new biological mechanisms that govern platelet function that may be targets in the prevention and treatment of cardiovascular disease.

Genetic factors play a well-documented role in platelet biology. Investigators have identified causal genetic variants of rare, inherited bleeding and clotting disorders, including Bernard-Soulier Syndrome (MIM: 231200), Glanzmann's Thrombasthenia (MIM: 273800), and Wiskott-Aldrich Syndrome (MIM: 301000). On a population scale, heritability studies demonstrated a strong genetic contribution to inter-individual platelet aggregation variance(1–3). Genome-wide association studies (GWAS) and other human genetic studies have identified numerous genes associated with population variation in platelet reactivity,

including *PEAR1*, *GP6*, *ADRA2A*, *MRVI1*, *JMJD1C*, and *SVIL*(6–10). Identification of associated genes has allowed for the further dissection of the mechanisms underlying platelet function using human, animal, and cellular systems(10–15).

However, GWAS and other similar association strategies typically only interrogate the association of common genetic variation (i.e., minor allele frequency [MAF] > 5%). Rare, coding genetic variants may also contribute to platelet traits in the population(16–18). With the advent of newer genetic technologies and analytical methods, investigators can now specifically examine the contribution of rare coding genetic variation through whole exome sequencing and exome chip genotyping. Several studies have applied such methods to platelet count and mean platelet volume, identifying previously unknown associations with *IQGAP2*, *MAP1A*, *MPL, FCER1A*, *JAK2*, and *TUBB1*, among others(16–18). Such studies of rare variants have not been performed for platelet reactivity traits in population studies, due to the relative scarcity of platelet reactivity measurements in larger cohort studies. Here, we report the results of a whole exome sequencing effort in the Framingham Heart Study (FHS) to identify rare variants that contribute to population variation in platelet reactivity. Our goal was to identify new genes associated with platelet aggregation not previously detected by GWAS and other human genetic association studies.

Methods

Participants

The participants in this investigation were from the Offspring cohort of the FHS. As our aim was to examine normal population variation in platelet reactivity, we excluded one individual with lymphoma. All other participants did not have other relevant conditions (e.g., hereditary anemias, end-stage kidney disease, and cirrhosis) or previously underwent relevant procedures (e.g., bone marrow transplant and splenectomy). Our initial analysis did not exclude those individuals with evidence for aspirin effects in order to maximize our discovery sample size. In total, our maximum sample size was 1,211 individuals. In secondary analyses, we repeated analyses excluding individuals with highly reduced responses to 5 mg/mL arachidonic acid (Supplemental Table 1). All participants completed informed consent procedures, and all experimental protocols were approved by the Boston University Institution Review Board.

Platelet Reactivity Measurements

Platelet reactivity was measured during Exam 5 of the FHS Offspring cohort as previously described(9). Platelet-rich plasma (PRP) was isolated from whole blood samples from fasting participants. We measured platelet reactivity using a four-channel aggregometer (Bio/Data Corporation) and three different agonists: ADP, epinephrine, and collagen (Bio/Data Corporation). We assessed aggregation up to 4 minutes after administration of ADP (0.05, 0.1, 0.5, 1.0, 3.0, 5.0, 10.0, and 15.0 μM) and up to 4 minutes after administration to epinephrine (0.01, 0.03, 0.05, 0.1, 0.5, 1.0, 3.0, 5.0, and 10.0 μM). The minimal concentration of ADP and epinephrine to produce a >50% aggregation response, termed threshold concentration (EC_{50}), was determined from these assays. We also analyzed maximal percent aggregation to two doses of ADP (3.0 μM and 5.0 μM) and of epinephrine

(1.0 μM and 3.0 μM). Finally, the lag time to aggregation to 190 μg/mL calf skin-derived type I collagen was assessed (Bio/Data Corporation). Aggregation testing was not conducted at higher agonist concentrations if a >50% aggregation response was observed, leading to differences in sample size among the traits analyzed (Table 1).

Whole Exome Sequencing

Whole exome sequencing of FHS samples in Freeze 5 was completed as part of a collaborative sequencing effort by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. In addition to the 1,702 FHS samples sequenced, 1,751 subjects from the Cardiovascular Health Study (CHS), and 10,990 subjects from the Atherosclerosis Risk in Communities Study (ARIC) were also sequenced in the same effort(19). However, the CHS and ARIC studies did not measure platelet reactivity and are not included in the current investigation.

DNA samples were constructed into Illumina paired-end libraries according to the manufacturer's protocol. Library construction protocols and oligonucleotide sequences used are freely available from the Baylor College of Medicine Human Genome Sequencing Center (www.hgsc.bcm.edu/content/protocols-sequencing-library-construction). Two, four, or six pre-capture libraries were pooled together and hybridized to the HGSC VCRome 2.1 design (42Mb, NimbleGen)(20). Whole exome sequencing was performed in paired-end mode using a single lane on the Illumina HiSeq 2000 or the HiSeq 2500 platform. Sequenced data were analyzed using the HGSC Mercury analysis pipeline (www.hgsc.bcm.edu/content/mercury). First, we de-multiplexed pooled samples with consensus assessment of sequence and variation (CASAVA). Then, we produced Binary Alignment/Map (BAM) files of reads mapped to the human reference sequence (GRCh37) using Burrows-Wheeler Alignment (BWA, bio-bwa.sourceforge.net/)(21;22). Aligned reads were recalibrated, sorted, marked for read duplications, and realigned near insertions/deletions (indels) using Genome Analysis ToolKit (GATK, http://www.broadinstitute.org/gatk/)(23). Single nucleotide variants (SNVs) and indels were called using the Atlas2 suite and stored in variant call files (VCF)(24;25). Mean depth of coverage among FHS samples was 84 fold for targeted regions.

Genotype calling, variant quality control, and variant filtering were completed in the three sequenced cohorts jointly. To produce a list of high-quality variants, all SNV calls were filtered on the following: low SNV posterior probability (<0.95), low variant read count (<3), variant read ratio <0.25 or >0.75, strand-bias of more than 99% variant reads in a single strand direction, and total coverage less than 10-fold. All SNV calls failing any of the filters, as well as reference calls with less than 10-fold coverage, were set to missing. Indel call filters were identical to those for SNV calls, except that a total coverage less than 30-fold was used for variant sites. We performed several variant-level quality control steps that excluded variants outside exon capture regions (VCRome 2.1), monomorphic, missing rate >20%, mappability score <0.8, and mean depth of coverage >500-fold. We also excluded variants not meeting Hardy-Weinberg equilibrium ($p<5\times10^{-6}$) in ancestry-specific groups. Samples were excluded if they had missingness >20%, or if they fell more than 6 standard

deviations (SD) in the FHS samples for mean depth, singleton count, heterozygote to homozygote ratio, or transition to transversion ratio.

Whole Exome Sequencing Statistical Methods

We examined seven platelet reactivity traits: 1) EC_{50} for ADP (ADP EC_{50}), 2) EC_{50} for epinephrine (Epi EC_{50}), 3) maximal percent aggregation to 3.0 μM ADP (MaxADP3), 4) maximal percent aggregation to 5.0 μM ADP (MaxADP5), 5) maximal percent aggregation to 1.0 μM epinephrine (MaxEpi1), 6) maximal percent aggregation to 3.0 μM epinephrine (MaxEpi3), and 7) collagen lag time (Table 1). These seven traits were the same as those examined in our prior GWAS of platelet reactivity(9). Examination of three different agonists enabled us to interrogate distinct platelet activation pathways and their responses to different agonist concentrations. After evaluation of distributions, we applied $\log_{10}$ transformations to three platelet reactivity traits (ADP EC_{50}, Epi EC_{50}, and collagen lag time). We then calculated residuals for each trait from linear regression models that adjusted for age, sex, and 10 principal components calculated from the FHS SNP Health Association Resource (SHARe) GWAS genotype data (Affymetrix 550K). Finally, we applied the rank-based inverse normal transformation to the resulting residuals. Variants were annotated using dbNSFP(26), and genes were defined by the UCSC gene model.

Single variant and gene-based associations were performed using RVFam which accounts for relatedness within our sample when performing association tests(27). A Bonferroni corrected significance threshold for number of single variants tested was $p=1.44\times10^{-7}$ (0.05 / 437,129 variants tested). For gene-based association tests, we employed the sum of squares (SSQ), threshold (T), and Madsen-Browning (MB) methods in RVFam using a variant MAF inclusion threshold of 5%(28–30). Performing multiple gene-based tests allowed us to examine associations in situations where variants included in the burden test had effects in the same (T/MB) and different (SSQ) directions. The gene-based Bonferroni corrected significance threshold was $p=2.35\times10^{-6}$ (0.05 / 21,301 genes tested). Following the gene-based association tests, we conducted leave-one-out analyses using our top gene-based findings to see whether the signals were driven by a single rare variant. Then, we evaluated the predicted deleteriousness of single variants from significant gene-based tests using Combined Annotation Dependent Depletion (CADD), with a scaled CADD score of 20 (corresponding to the top 1% deleterious variants in the genome as indicated by the authors) as a threshold for predicted deleteriousness(31). Protein sequences for HYAL2 in humans and orthologs were obtained from UniProt(32), and alignment was assessed with Clustal Omega(33) under default settings. We also queried the expression levels in platelets of the most strongly associated genes using previously reported and publicly available RNA-seq data from 32 platelet samples (GEO Accession: GSE65705) to determine whether these genes are expressed in platelets(34).

Post-hoc Power Calculations

Power analysis was conducted using simulations based on the FHS sample. SOLAR software was used to simulate phenotypes with a polygenic heritability 0.25 for the random effects and the quantitative trait locus (QTL) was set to have minor allele frequency of 0.001, 0.01, 0.05, 0.1, 0.2, and 0.3, and QTL variance of 0.005, 0.01, 0.02 and 0.025. We

used the same samples that had MaxEpi3 (n=511), MaxADP5 (n=779), collagen (n=1026) and MaxADP3 (n=1211) as given in Table 1 to test association between phenotype and QTL. The lmekin function from coxme R package that uses a linear mixed effects model to account for familial correlation was used for association testing. 1,000 replicates were used to evaluate power with significance threshold of 1.4E-7.

GeneSTAR Whole Genome Sequencing Replication

In an attempt to gain more evidence supporting our associations, we analyzed data from the Genetic Study of Atherosclerosis Risk (GeneSTAR). GeneSTAR is an ongoing prospective study designed to determine environmental, phenotypic, and genetic causes of premature cardiovascular disease. Participants were identified from probands with a premature coronary disease event prior to 60 years of age who were identified at the time of hospitalization in any of 10 Baltimore area hospitals. Apparently healthy siblings of the probands and offspring of the siblings and probands were screened for traditional coronary disease and stroke risk factors as part of a 2 week trial of platelet function prior to and following aspirin therapy, 81 mg/day, from 2003–2006. Participants were asked to refrain from aspirin use for 2 weeks prior to baseline assessment, and all measures described here were obtained prior to the commencement of aspirin therapy. Exclusion criteria included current or prior relevant medical conditions (e.g., vascular thrombotic event, bleeding disorders, and menorrhagia) and use of medications that may interfere with measured outcomes (e.g., anti-platelet, anti-coagulant, and nonsteroidal anti-inflammatory agents). All study participants provided written informed consent prior to participation.

Platelet aggregation measures are described in detail elsewhere(9). In brief, platelet-rich plasma (PRP) was isolated from fasting blood samples and measured platelet aggregation after addition of agonists using a four-channel aggregometer. Maximal aggregation (% aggregation) was recorded for periods of 5 minutes after 2.0 μM ADP and 5 minutes after 2.0 μM epinephrine administration.

Genomes were sequenced with >30x coverage on the Illumina HiSeq platform at Illumina, Inc (San Diego, CA). The full details of the whole genome sequencing and variant calling pipeline are described in detail elsewhere(35). Briefly, assembly of each individual genome was performed using the CASAVA package. CASAVA then calls the genomic consensus sequence using a probabilistic algorithm and compares it to the reference sequence to identify homozygous or heterozygous SNPs. Data processing to generate a multi-sample VCF file for each chromosome from the Illumina single-sample SNP VCF files provided in Illumina's standard deliverable package was performed locally at Johns Hopkins University. The individual VCF files were merged using VCFtools. Using custom scripts, the multi-sample VCF files were backfilled to include homozygous reference genotypes and depth of coverage from the sites.txt files. Custom QC scripts confirmed that the multi-sample VCFs and the single-sample VCFs had the same number of heterozygous and homozygous alternate genotypes. VCFtools was used to confirm that all subjects were included in each multi-sample VCF. Variants with genotype quality (GQ) < 20 or depth (DP) < 7 were filtered. Additionally, variants that fell into regions of segmental duplication were removed.

This study sample included 126 African American and 156 European American individuals with both platelet function data and whole genome sequencing data (Supplemental Table 2). Inverse normal transformed residuals of maximal aggregation to 2.0 *u*M ADP and 2.0 *u*M epinephrine were obtained to match similar dose profiles to the discovery analysis. The residuals were analyzed for genetic association using RVFam that accounts for familial correlation.

Results

No single variant showed an association with platelet reactivity traits that surpassed the correction for multiple testing (Supplemental Table 3). Lack of a significant single variant association is likely due to limited sample size, as well as low coverage of potential regulatory variation in the genome with an exome-sequencing based approach. However, single nucleotide polymorphisms (SNPs) in several known (e.g., *MRVI1* with ADP and *GP6* with collagen) and potentially novel (e.g., *ANPEP*, *HIVEP3*, *NEXN*, and *DISP1*) loci showed suggestive associations with at least one platelet reactivity trait ($p<1.0\times10^{-5}$) (Supplemental Table 3). The associated SNPs in the known *MRVI1* and *GP6* loci were common variants (MAF > 5%). The lone associated *MRVI1* variant was intronic near an exon-intron junction, while the five associated coding variants in *GP6* included 3 synonymous and 2 predicted benign missense variants, A249T and P219S. The associated *MRVI1* and *GP6* variants were in high linkage disequilibrium with lead SNPs—$r^2=0.558$ with rs7940646 and $r^2>0.760$ with rs1671152, respectively, as calculated using the CEU population from 1000 Genomes in SNAP—from previous platelet aggregation GWAS(9;36).

Despite a lack of significant single variant results, we did observe two gene-based associations that passed correction for multiple testing: *HYAL2* with MaxADP3 ($p=1.07\times10^{-7}$) and *GSTZ1* with MaxEpi3 ($p=1.62\times10^{-6}$) (Table 2). *HYAL2* also showed suggestive association with epinephrine-induced aggregation in the same direction of effect as the ADP association ($p=2.64\times10^{-5}$) (Supplemental Table 4, Supplemental Table 5a). The *HYAL2* gene-based association was comprised of 12 rare variants (MAF<0.001), with 5 individually showing nominal (p<0.05) positive (i.e., increased reactivity) association with MaxADP3 (Figure 1, Table 3). Carriers of these 12 rare variants tended to show increased platelet reactivity to both ADP and epinephrine (Figure 2a). Exclusion of individual variants from the gene-based test did not abrogate the association signaling, indicating the *HYAL2* gene-based association was not driven by a single variant (Excl. p in Table 3). Likewise, excluding individuals with evidence for aspirin effect did not dramatically affect the results (Supplemental Tables 6–7). The four most strongly associated *HYAL2* rare variants were predicted to be deleterious by CADD (Table 3). Each of the nominally associated *HYAL2* rare variants occurred at conserved amino acid residues (Supplemental Figure 1). A previously undescribed nonsense variant (Q406*) had the most deleteriousness predicted effects and removes a glycophosphatidylinositol (GPI) anchor from the HYAL2 protein. Additionally, another variant contributing to the *HYAL2* gene-based association was a missense variant at a known glycosylation site (N357S) (Figure 1). Using NetNGlyc 1.0 (www.cbs.dtu.dk/services/NetNGlyc), the replacement of this serine by an asparagine removes this predicted glycosylation residue.

The other significant gene-based association of *GSTZ1* with MaxEPI1 ($p=1.62\times10^{-6}$) was comprised of three variants (Table 2, Table 4). Each variant was seen in one individual and nominally associated with increased epinephrine-induced aggregation. Similar to *HYAL2*'s association with aggregation to multiple agonists, we also observed a suggestive association of *GSTZ1* with MaxADP5 ($p=1.78\times10^{-5}$) (Supplemental Table 4, Supplemental Table 5b) and that carriers of these rare variants in GSTZ1 tended to have increased reactivity to ADP and epinephrine (Figure 2b). Other suggestive gene-based associations included *MAPRE1* with ADP EC_{50} ($p=7.26\times10^{-6}$) and *AR* with MaxADP3 ($p=7.39\times10^{-6}$) (Table 2, Supplemental Table 5c–5d). Excluding individuals with potential aspirin effects altered these results very little (Supplemental Tables 6). Significant (*HYAL2, GSTZ1*) and suggestive (*AR, MAPRE1*) gene-based associations were identified with the Madsen-Browning test and/or Threshold test. The SSQ test which is most powerful when individual variants within genes have associations in both directions, similar to the commonly applied SKAT test, did not reveal any gene-based associations with $p<1.0\times10^{-5}$. Nonetheless, we further queried these gene-based association results for candidate genes identified through previous GWAS of platelet aggregation and studies of inherited platelet aggregation disorders (Supplemental Table 9a). Although none of these associations approached our corrected threshold for statistical significance, there were suggestive associations with *ITGA2* and *PIK3CG* (Supplemental Table 9b).

Using whole genome sequencing data in GeneSTAR African American (AA) and European ancestry (EA) samples (n<160), we examined the association of our top gene-based associations (*HYAL2, GSTZ1, MAPRE1*) with maximal aggregation to 2*u*M ADP or 2*u*M epinephrine. These doses were selected as the closest ones available to match the discovery analysis doses in FHS. Analysis of variants on the X-chromosome (i.e., *AR*) was not available. We found moderate evidence supporting the relationship of *GSTZ1* with MaxEpi2*u*M (SSQ test p-value = 0.047) and *MAPRE1* with MaxADP2*u*M (MB/T tests p-value = 0.023) in the GeneSTAR AA sample (Supplemental Table 10). The opposite betas of *MAPRE1* associations in FHS ADP EC_{50} and GeneSTAR MaxADP 2*u*M are expected as ADP EC_{50} is negatively correlated with MaxADP. In the GeneSTAR EA sample, we did observe associations with p<0.05. but the direction of effect was opposite to those in FHS (Supplemental Table 10).

To further interrogate the relevance of the most strongly associated genes, we queried the expression levels of these genes in platelets using existing RNA-seq data from 32 platelet samples (Supplemental Table 11)(34). Each of the genes from Table 2 and Supplemental Table 3 were expressed in platelets. However, some transcripts (e.g., MAPRE1, ANPEP, GP6, and NEXN) were more highly expressed than others which demonstrated moderate expression (e.g., HYAL2, GSTZ1, and AR).

Discussion

In this investigation, we performed a whole exome sequencing study of platelet aggregation using population-based cohort data. To our knowledge, whole exome sequencing investigations of platelet traits have mainly focused on rare inherited platelet bleeding and platelet disorders(37). Our present study allowed us to examine the contribution of rare

coding variants to population variation in platelet aggregation. By prioritizing rare variants using gene-based tests, we identified novel associations not previously detected by GWAS. Specifically, we found associations of rare coding variation in *HYAL2* and *GSTZ1* with platelet reactivity to ADP and epinephrine. Grouping rare variants by gene units allowed us to observe associations we were underpowered to detect when only examining single variants as shown in previous studies of blood traits(38). By doing so, we not only gain insight into the biology underlying platelets' aggregation response to agonists, but show that grouping variants by functional annotations can be fruitful as whole exome and whole genome sequencing in large cohorts becomes more common.

The strongest association in our whole exome sequencing study was between ADP-induced aggregation and rare variation in *HYAL2*. Gene-based tests for *HYAL2* were also associated with increased epinephrine-induced aggregation (MaxEpi1, $p=2.64\times10^{-5}$), indicating that *HYAL2* may be associated with overall platelet reactivity, independent and downstream of the initial agonist activation. Animal knockouts of *HYAL2* have demonstrated its effects on platelets and red blood cells(39;40). HYAL2 functions in the extracellular degradation of hyaluronan in acidic environments, with inactivation of HYAL2 leading to increased plasma hyaluronan levels(39). The regulation and catalysis of hyaluronan at injury sites is an important mediator of local inflammation and wound healing, primary functions of platelets. Further studies have shown that HYAL2 is transported from alpha-granules to the cell surface during platelet activation to degrade hyaluronan into inflammatory, angiogenic, and hemostatic signaling fragments(41;42). Together with the evidence from the literature, our study suggests that HYAL2 modulates platelet reactivity in humans and that its degradation of hyaluronan may be an important mediatory process. Thus, HYAL2 levels, its activity, or hyaluronan measures could be potential biomarkers for abnormal wound healing or platelet function.

Twelve rare variants (MAF<0.01) were combined in generating the gene-based test for *HYAL2* (Figure 1, Table 3). In the literature, many gene-based signals are driven by the association of a single variant. That does not appear to be the case with the *HYAL2* association here as five single variants showed associations ($p<0.05$) and systemically excluding individual variants did not abrogate the association (Table 3). Of these five associated rare variants in *HYAL2*, four were predicted to be deleterious by CADD(31) and give rise to amino acid changes in the second half of the protein. One of these variants, rs139202918, results in the substitution of a serine for an asparagine (N357S) at a predicted glycosylation residue. Consequently, this amino acid substitution is predicted to remove this glycosylation residue at position 357. Studies of protein family member HYAL1 have shown that N-glycosylation is required for proper secretion and enzymatic activity(43). Homology between HYAL1 and HYAL2 N-glycosylation sites at positions 350 and 357, respectively, is high and suggests that HYAL2 N-glycosylation is important for proper protein trafficking and function. A second variant, not previously observed in dbSNP or in Exome Aggregation Consortium (ExAC) data, introduces a stop codon at position 406 (Q406*) (http://exac.broadinstitute.org [March 2016])(44). This nonsense mutation removes a GPI anchor from the HYAL2 protein sequence (Figure 1). These rare variants that alter N-glycosylation or the addition of a functional GPI anchor may prevent proper trafficking to the cell membrane during the initiation of platelet activation. As the gene-based tests showed an

association of *HYAL2* rare variants with increased reactivity to ADP and epinephrine, the predicted inability of HYAL2 to be properly trafficked to the cell membrane may prevent the catalysis of high molecular weight hyaluronan, prohibit the cessation of platelet activation, and propagate the platelets' aggregation response. The specific functional consequences of the other implicated variants in *HYAL2* are unclear; however, these deleterious variants may impede the catalytic activity of HYAL2 in degrading high molecular weight hyaluronan and allow platelet aggregation to propagate.

In addition to the association of *HYAL2* with platelet aggregation, we also found association of *GSTZ1* with increased ADP- and epinephrine- induced aggregation. GSTZ1 functions in the tyrosine/phenylalanine degradation and the degradation of several toxic compounds. *Gstz* −/− mice do not exhibit abnormalities in erythrocyte or platelet indices, but do demonstrate reductions in leukocyte measures(45). There were no notable white blood cell phenotypes in the *GSTZ1* rare variant carriers in FHS. Nonetheless, leukocyte count has been associated with platelet reactivity, further indicating the important cross-talk between platelets and leukocytes during thrombosis and wound healing(46). Our study suggests that rare variants in *GSTZ1* modulate general platelet reactivity. Although *GSTZ1* is expressed in platelets, whether *GSTZ1* acts through platelets, leukocytes, or both remains a question to be examined in future studies.

Finally, we observed several suggestive single variant and gene-based associations, including *AR* and *MAPRE1* with decreased ADP-induced aggregation. A common missense variant in *NEXN*, rs1166698 (Gly245Arg), was associated with longer lag times to collagen. NEXN is expressed in cardiovascular tissues including atria and vessels, and involved in cell adhesion. Previous studies have linked variants in the same exon as rs1166698 to dilated cardiomyopathy and heart failure(47). The association with *AR* was especially intriguing due to the known sex differences in platelet traits(48–50). Observing association of platelet reactivity with genetic variation in *AR* as previously suggested indicates that these sex differences may result from differences in AR-mediated transcriptional regulation(51). MAPRE1, also known as EB1, functions in microtubule stability and assembly, two essential processes in platelet activation and aggregation. In fact, other studies have used EB1 antibodies to visualize microtubule dynamics in resting and activated platelets(52). Rare variants that impair MAPRE1's function in microtubule dynamics may result in decreased platelet reactivity to presented agonists. The roles of *NEXN*, *GSTZ1*, *AR*, and *MAPRE1* in platelet aggregation will need to be further explored in future functional studies.

Three of our most strongly associated genes—*HYAL2*, *GSTZ1*, and *AR*—showed low to moderate RNA expression in platelets (average RPKM of 2.09, 3.18, and 1.92, respectively) (Supplemental Table 11). Interpretation of platelet RNA expression is complicated as active genomic transcription is only completed in the precursor megakaryocyte cells. The translation of these RNAs into proteins (e.g., the production of HYAL2 protein for storage in alpha-granules) or the functional implication of these proteins (e.g., the effects of AR on megakaryocyte gene transcription) may have already occurred. Additionally, it is possible the genes could have an impact on platelet function via non-megakaryocyte and non-platelet mechanisms. Future studies utilizing RNA and protein expression in precursor cells and

mature platelets are needed to disentangle temporal and mechanistic effects on platelet aggregation.

We found limited support for significant associations of *GSTZ1, MAPRE1* and *HYAL2* with ADP or epinephrine induced platelet aggregation in an independent study, the GeneSTAR study. These findings were relatively weak and in some cases in opposite effect directions. Replicating rare variants for rare traits is challenging(53) particularly with a small sample size. Even when rare variants are observed in the target genes, the bulk of variants observed often will not overlap those in the discovery sample. Coding variants in the replication sample may localize to domains that are not critical for protein function, and/or could be benign in their functional effects. Thus, much larger samples or orthogonal methods in humans and model systems may be required to provide a high level of evidence.

Our whole exome sequencing study of platelet reactivity is subject to several limitations. Examining rare variants, even by gene-based tests, requires substantial sample size to detect effects. Our study with a maximum sample size of 1,211 lacks sufficient power to detect many associations (Supplemental Table 12), particularly with that of rare variants and of common variants of smaller effect that were detected in previous GWAS. Unlike platelet counts, platelet function tests remain a relatively scarce phenotype in larger cohort studies, with FHS being among the largest in the world. Nonetheless, the need remains to replicate our associations further in independent cohorts—available whole genome sequencing in the GeneSTAR study currently encompassed a small sample which will be expanded upon in the future. Our platelet function assays also cover only a subset of known activation pathways. Namely, we are missing primary data on other important modalities like thrombin-based activation, shear stress, and thromboxane pathways. Despite this, the FHS sample still represents a wide range of experimental conditions in a large population-based study. Sample size also limited the types of rare variants to specifically examine in our gene-based association tests. As opposed to limiting rare variants to those with high predicted deleteriousness, we included all exonic variants with MAF < 0.05—including those with low damaging potential—to increase the number of variants tested. However, we do include annotation and interpretation of variant deleteriousness of our strongest associations. Finally, we do not present functional evidence for the individual variants identified in the associated genes identified in our burden tests. Currently, the functional consequences of these variants are inferred from several annotation tools and the current literature. Future experiments are needed in order to show the functional implications of the variants in *HYAL2* and *GSTZ1.* Even with these study limitations, the reported associations add significantly to collective knowledge on the genetics of platelet reactivity and may aid in future human, animal, and cellular interrogations of platelet functions.

Using whole exome sequencing in the FHS, we identified gene-based associations of rare variants in *HYAL2* and *GSTZ1* with platelet aggregation. While animal and cellular based studies have previously shown *HYAL2* to have a role in hematological processes, our study is the first to show a relationship of *HYAL2* with platelet traits in humans. By using whole exome sequencing, we are able to propose functional variants for *HYAL2* that may prevent proper trafficking of the protein to the membrane, a topic for future investigations. We are also able to identify possible associations of new platelet reactivity genetic candidates

GSTZ1, *AR*, and *MAPRE1*. More broadly, our study is an example of grouping rare variants into functional bins to detect novel associations that would be missed by conventional single variant association analyses. As whole genome sequencing in larger and larger population-based cohorts becomes feasible, the use of functional annotations to group rare variation will become more important for the detection of novel associations and more importantly their mechanistic and clinical implications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank all individuals for their participation in this study. 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. 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), and for quality control by Framingham Heart Study investigators using genotypes in the SNP Health Association Resource (SHARe) project. The computational work reported on in this paper was performed on the Shared Computing Cluster, administered by Boston University's Research Computing Services (www.bu.edu/tech/support/research/) and on the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. Funding support for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE Consortium" was provided by the NIH through the American Recovery and Reinvestment Act of (5RC2HL102419). Data for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE Consortium" were provided by Eric Boerwinkle (ARIC), L. Adrienne Cupples (FHS), and Bruce Psaty (CHS). Sequencing was carried out at the Baylor Genome Center (U54 HG003273). The authors would like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at http://exac.broadinstitute.org/about.

GeneSTAR was funded by grants from the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518, HL087698, HL112064) 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.

Reference List

(1). O'Donnell CJ, Larson MG, Feng D, Sutherland PA, Lindpaintner K, Myers RH, et al. Genetic and environmental contributions to platelet aggregation: the Framingham heart study. Circulation 2001;103:3051–6. [PubMed: 11425767]

(2). Bray PF, Mathias RA, Faraday N, Yanek LR, Fallin MD, Herrera-Galeano JE, et al. Heritability of platelet function in families with premature coronary artery disease. J Thromb Haemost 2007;5:1617–23. [PubMed: 17663734]

(3). Johnson AD. The genetics of common variation affecting platelet development, function and pharmaceutical targeting. J Thromb Haemost 2011;9 Suppl 1:246–57. [PubMed: 21781261]

(4). Davi G, Patrono C. Platelet activation and atherothrombosis. N Engl J Med 2007;357:2482–94. [PubMed: 18077812]

(5). Hennekens CH, Dyken ML, Fuster V. Aspirin as a therapeutic agent in cardiovascular disease: a statement for healthcare professionals from the American Heart Association. Circulation 1997;96:2751–3. [PubMed: 9355934]

(6). Qayyum R, Becker LC, Becker DM, Faraday N, Yanek LR, Leal SM, et al. Genome-wide association study of platelet aggregation in African Americans. BMC Genet 2015;16:58. [PubMed: 26024889]

(7). Qayyum R, Snively BM, Ziv E, Nalls MA, Liu Y, Tang W, et al. A meta-analysis and genome-wide association study of platelet count and mean platelet volume in african americans. PLoS Genet 2012;8:e1002491. [PubMed: 22423221]

(8). Lewis JP, Ryan K, O'Connell JR, Horenstein RB, Damcott CM, Gibson Q, et al. Genetic variation in PEAR1 is associated with platelet aggregation and cardiovascular outcomes. Circ Cardiovasc Genet 2013;6:184–92. [PubMed: 23392654]

(9). Johnson AD, Yanek LR, Chen MH, Faraday N, Larson MG, Tofler G, et al. Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. Nat Genet 2010;42:608–13. [PubMed: 20526338]

(10). Edelstein LC, Luna EJ, Gibson IB, Bray M, Jin Y, Kondkar A, et al. Human genome-wide association and mouse knockout approaches identify platelet supervillin as an inhibitor of thrombus formation under shear stress. Circulation 2012;125:2762–71. [PubMed: 22550155]

(11). Kauskot A, Vandenbriele C, Louwette S, Gijsbers R, Tousseyn T, Freson K, et al. PEAR1 attenuates megakaryopoiesis via control of the PI3K/PTEN pathway. Blood 2013;121:5208–17. [PubMed: 23667054]

(12). Kauskot A, Di MM, Loyen S, Freson K, Verhamme P, Hoylaerts MF. A novel mechanism of sustained platelet alphaIIbbeta3 activation via PEAR1. Blood 2012;119:4056–65. [PubMed: 22371881]

(13). Sun Y, Vandenbriele C, Kauskot A, Verhamme P, Hoylaerts MF, Wright GJ. A Human Platelet Receptor Protein Microarray Identifies the High Affinity Immunoglobulin E Receptor Subunit alpha (FcepsilonR1alpha) as an Activating Platelet Endothelium Aggregation Receptor 1 (PEAR1) Ligand. Mol Cell Proteomics 2015;14:1265–74. [PubMed: 25713122]

(14). Vandenbriele C, Kauskot A, Vandersmissen I, Criel M, Geenens R, Craps S, et al. Platelet endothelial aggregation receptor-1: a novel modifier of neoangiogenesis. Cardiovasc Res 2015;108:124–38. [PubMed: 26156496]

(15). Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, Pistis G, et al. New gene functions in megakaryopoiesis and platelet formation. Nature 2011;480:201–8. [PubMed: 22139419]

(16). Eicher JD, Chami N, Kacprowski T, Normura A. Platelet-related variants identified by exome chip meta-analysis in 157,293 individuals. Am J Hum Genet 2016.

(17). Auer PL, Johnsen JM, Johnson AD, Logsdon BA, Lange LA, Nalls MA, et al. Imputation of exome sequence variants into population- based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO Exome Sequencing Project. Am J Hum Genet 2012;91:794–808. [PubMed: 23103231]

(18). Auer PL, Teumer A, Schick U, O'Shaughnessy A, Lo KS, Chami N, et al. Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. Nat Genet 2014;46:629–34. [PubMed: 24777453]

(19). Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet 2009;2:73–80. [PubMed: 20031568]

(20). Bainbridge MN, Wang M, Wu Y, Newsham I, Muzny DM, Jefferies JL, et al. Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol 2011;12:R68. [PubMed: 21787409]

(21). Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60. [PubMed: 19451168]

(22). Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9. [PubMed: 19505943]

(23). DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–8. [PubMed: 21478889]

(24). Challis D, Yu J, Evani US, Jackson AR, Paithankar S, Coarfa C, et al. An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics 2012;13:8. [PubMed: 22239737]

(25). Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156–8. [PubMed: 21653522]

(26). 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 2016;37:235–41. [PubMed: 26555599]

(27). Chen MH, Yang Q. RVFam: an R package for rare variant association analysis with family data. Bioinformatics 2016;32:624–6. [PubMed: 26508760]

(28). Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 2008;83:311–21. [PubMed: 18691683]

(29). Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet 2009;5:e1000384. [PubMed: 19214210]

(30). Pan W Asymptotic tests of association with multiple SNPs in linkage disequilibrium. Genet Epidemiol 2009;33:497–507. [PubMed: 19170135]

(31). Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5. [PubMed: 24487276]

(32). UniProt: a hub for protein information. Nucleic Acids Res 2015;43:D204–D212. [PubMed: 25348405]

(33). Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res 2015;43:W580–W584. [PubMed: 25845596]

(34). Eicher JD, Wakabayashi Y, Vitseva O, Esa N, Yang Y, Zhu J, et al. Characterization of the platelet transcriptome by RNA sequencing in patients with acute myocardial infarction. Platelets 2016;27:230–9. [PubMed: 26367242]

(35). Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. Nat Commun 2016;7:12522. [PubMed: 27725671]

(36). Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 2008;24:2938–9. [PubMed: 18974171]

(37). Maclachlan A, Watson SP, Morgan NV. Inherited platelet disorders: Insight from platelet genomics using next-generation sequencing. Platelets 2016;1–6. [PubMed: 26695713]

(38). Chami N, Chen MH, Slater AJ, et al. Several new pleiotropic variants associated with red blood cell traits identified by exome genotyping. Am J Hum Genet 2016.

(39). Jadin L, Wu X, Ding H, Frost GI, Onclinx C, Triggs-Raine B, et al. Skeletal and hematological anomalies in HYAL2-deficient mice: a second type of mucopolysaccharidosis IX? FASEB J 2008;22:4316–26. [PubMed: 18772348]

(40). Onclinx C, Dogne S, Jadin L, Andris F, Grandfils C, Jouret F, et al. Deficiency in mouse hyaluronidase 2: a new mechanism of chronic thrombotic microangiopathy. Haematologica 2015;100:1023–30. [PubMed: 25934767]

(41). Albeiroti S, Ayasoufi K, Hill DR, Shen B, de la Motte CA. Platelet hyaluronidase-2: an enzyme that translocates to the surface upon activation to function in extracellular matrix degradation. Blood 2015;125:1460–9. [PubMed: 25411425]

(42). de la Motte C, Nigro J, Vasanji A, Rho H, Kessler S, Bandyopadhyay S, et al. Platelet-derived hyaluronidase 2 cleaves hyaluronan into fragments that trigger monocyte-mediated production of proinflammatory cytokines. Am J Pathol 2009;174:2254–64. [PubMed: 19443707]

(43). Goto Y, Niwa Y, Suzuki T, Uematsu S, Dohmae N, Simizu S. N-glycosylation is required for secretion and enzymatic activity of human hyaluronidase1. FEBS Open Bio 2014;4:554–9.

(44). Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–91. [PubMed: 27535533]

(45). Lim CE, Matthaei KI, Blackburn AC, Davis RP, Dahlstrom JE, Koina ME, et al. Mice deficient in glutathione transferase zeta/maleylacetoacetate isomerase exhibit a range of pathological changes

and elevated expression of alpha, mu, and pi class glutathione transferases. Am J Pathol 2004;165:679–93. [PubMed: 15277241]

(46). Faraday N, Yanek LR, Vaidya D, Kral B, Qayyum R, Herrera-Galeano JE, et al. Leukocyte count is associated with increased platelet reactivity and diminished response to aspirin in healthy individuals with a family history of coronary artery disease. Thromb Res 2009;124:311–7. [PubMed: 19185906]

(47). Hassel D, Dahme T, Erdmann J, Meder B, Huge A, Stoll M, et al. Nexilin mutations destabilize cardiac Z-disks and lead to dilated cardiomyopathy. Nat Med 2009;15:1281–8. [PubMed: 19881492]

(48). Faraday N, Goldschmidt-Clermont PJ, Bray PF. Gender differences in platelet GPIIb-IIIa activation. Thromb Haemost 1997;77:748–54. [PubMed: 9134654]

(49). Becker DM, Segal J, Vaidya D, Yanek LR, Herrera-Galeano JE, Bray PF, et al. Sex differences in platelet reactivity and response to low-dose aspirin therapy. JAMA 2006;295:1420–7. [PubMed: 16551714]

(50). Kelton JG, Carter CJ, Santos A, Hirsh J. Sex related differences in platelet function: the effect of aspirin. Blood 1982;59:625–7. [PubMed: 7059670]

(51). Kuliczkowski W, Filus A, Kuliczkowska-Plaksej J, Trzmiel-Bira A, Laczmanski L, Karolko B, et al. Androgen receptor polymorphism and platelet reactivity in healthy men. Thromb Res 2010;126:e65–e67. [PubMed: 19786297]

(52). Patel-Hett S, Richardson JL, Schulze H, Drabek K, Isaac NA, Hoffmeister K, et al. Visualization of microtubule growth in living platelets reveals a dynamic marginal band with multiple microtubules. Blood 2008;111:4605–16. [PubMed: 18230754]

(53). Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. Genome Med 2015;7:16. [PubMed: 25709717]

What is known about this topic?

- Modulating platelet reactivity with anti-platelet medication is a common strategy for secondary prevention of cardiovascular disease. Identifying novel activation mechanisms can improve therapeutic strategies.
- Common and rare genetic factors contribute to platelet reactivity traits in the population. However, no population-based assessment of rare variants using whole exome sequencing has been performed.

What does this paper add?

- This study of platelet reactivity in the Framingham Heart Study is the first in a population-based cohort to identify rare variants from whole exome sequencing that contribute to platelet reactivity.
- Using gene-based association tests, we identify association of rare variation in *HYAL2* and *GSTZ1* with platelet reactivity in the population.
- Our data suggests that rare variants that disrupt the transport of HYAL2 to the platelet plasma membrane influence platelet aggregation response to agonists.

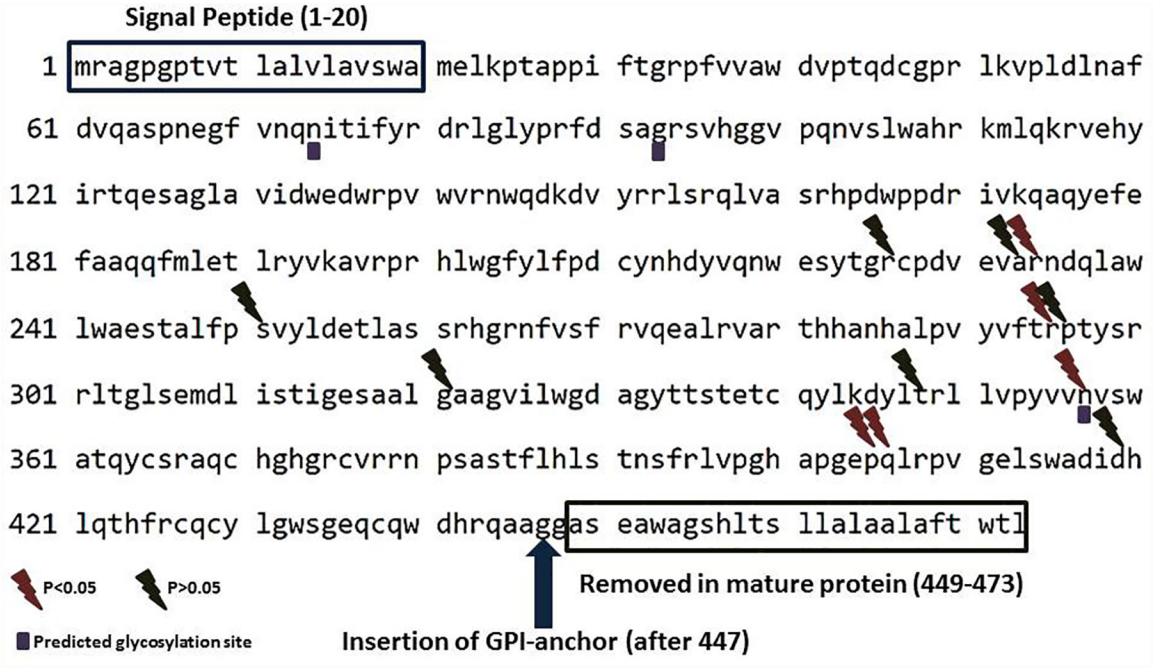

Figure 1: Single variants used in gene-based associations of *HYAL2*.

Twelve genetic variants were included in gene-based associations of *HYAL2* with ADP-induced reactivity (MaxADP3, $p=1.07\times10^{-7}$). Five genetic variants—denoted in red—showed individual associations with MaxADP3 ($p<0.05$). A nonsense variant (Q406*) removes a glycophosphatidylinositol (GPI) anchor from the resulting protein, while one missense variant (N357S) occurs at a predicted glycosylation site. Such associations indicate that improper HYAL2 trafficking and function influences platelet reactivity.

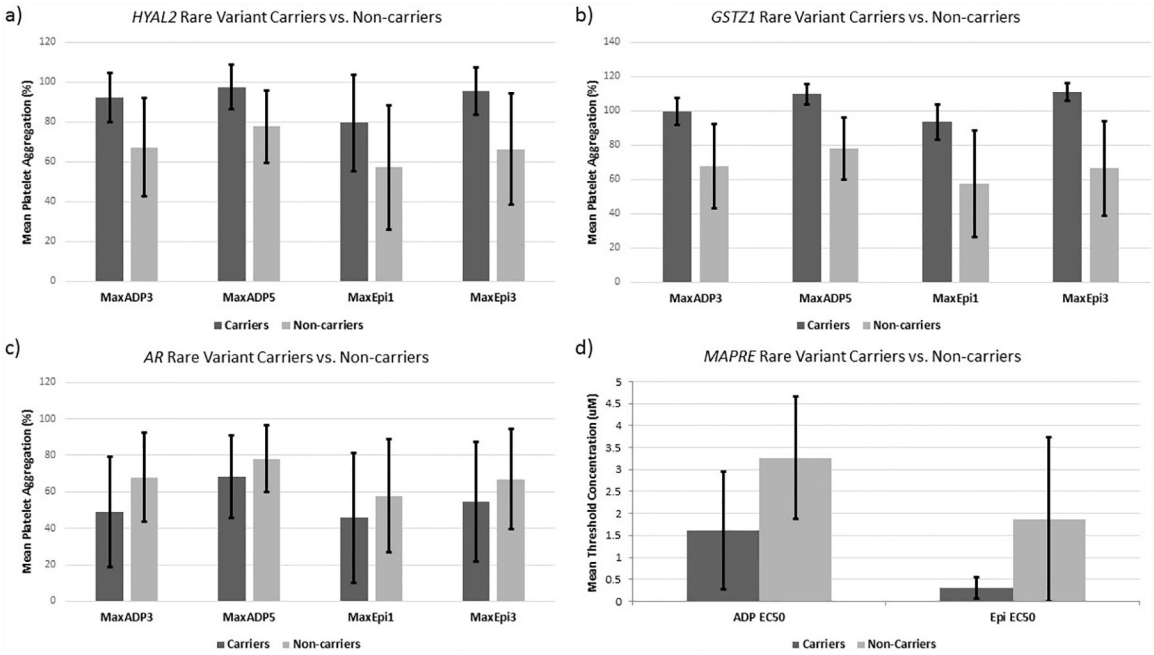

Figure 2: Mean platelet aggregation measures of rare variant carriers and non-carriers.

Comparing mean platelet aggregation measures of individuals carrying and not carrying the rare variants used in gene-based associations of (a) *HYAL2*, (b) *GSTZ1*, (c) *AR*, and (d) *MAPRE1*. Means expressed as percentage of maximal aggregation to 3 μM ADP (MaxADP3, 5 μM ADP (MaxADP5), 1 μM epinephrine (MaxEpi1), 3 μM epinephrine (MaxEpi3) and ADP or Epi EC50 (*u*M) with error bars representing standard deviation.

Table 1:

Summary of platelet reactivity phenotypes examined

Trait	N	Mean	SD	Description
ADP EC_{50} (μM)	1053	3.25	1.40	Threshold ADP concentration to produce a >50% aggregation response
Collagen (sec)	1026	81.34	19.88	Lag time to 190 μg/mL calf skin-derived type I collagen
Epi EC_{50} (μM)	1053	1.85	2.75	Threshold epinephrine concentration to produce a >50% aggregation response
MaxADP3 (%)	1211	67.66	24.69	Maximal percent aggregation to 3.0 μM ADP
MaxADP5 (%)	779	77.93	18.24	Maximal percent aggregation to 5.0 μM ADP
MaxEpi1 (%)	960	57.53	31.13	Maximal percent aggregation to 1.0 μM epinephrine
MaxEpi3 (%)	522	66.68	27.78	Maximal percent aggregation to 3.0 μM epinephrine

Table 2:

Summary of gene-based associations with platelet reactivity ($p<1.0\times10^{-5}$)

Madsen-Browning (MB) Test							
Trait	**Gene**	**Location**	**# SNPs**	**cMAF**	**Beta**	**SE**	**P-value**
MaxADP3 *	*HYAL2*	3p21.13	12	0.00663	0.00069	0.00013	1.07×10^{-7}
MaxEpi3 *	*GSTZ1*	14q24.3	3	0.00287	0.00259	0.00054	1.62×10^{-6}
ADP EC_{50}	*MAPRE1*	20q11.21	5	0.00427	−0.00117	0.00026	7.26×10^{-6}
MaxADP3	*AR*	Xq12	12	0.01277	−0.00053	0.00012	7.39×10^{-6}
Threshold (T) Test							
Trait	**Gene**	**Location**	**# SNPs**	**cMAF**	**Beta**	**SE**	**P-value**
MaxADP3 *	*HYAL2*	3p21.13	12	0.00663	1.317	0.251	1.58×10^{-7}
MaxEpi3 *	*GSTZ1*	14q24.3	3	0.00287	2.708	0.565	1.62×10^{-6}
MaxADP3	*AR*	Xq12	12	0.01277	−0.675	0.151	7.66×10^{-6}

* Association surpasses threshold for multiple testing ($p<2.15\times10^{-6}$)

Abbreviations: SNPs, single nucleotide polymorphisms; cMAF, combined minor allele frequency; SE, standard error; MaxADP3, Maximal percent aggregation to 3.0 μM ADP; MaxEpi3, Maximal percent aggregation to 3.0 μM epinephrine; ADP EC_{50}, Threshold ADP concentration to produce a >50% aggregation response

Table 3:

Association results of single variants in *HYAL2* used in gene-based tests for MaxADP3 (n=1,211)

Chr:pos	rsID	MAF	A1	A2	N Carriers	Beta	SE	P-value	Mutation	CADD	Excl. p[c]
3:50,355,725	rs202020190	0.000413	G	A	1	−0.07	1.00	0.941	D419D	3.755	1.72×10^{-8}
3:50,355,766	[a]	0.000413	G	A	1	2.14	1.00	0.0314	Q406*	37	1.17×10^{-6}
3:50,355,768	rs150262015	0.000413	G	A	1	2.55	1.00	0.0105	P405L	21.3	2.37×10^{-6}
3:50,355,912	rs139202918	0.000413	T	C	1	2.75	1.00	0.00566	N357S	25.5	3.36×10^{-6}
3:50,355,939	[a]	0.000826	G	A	2	0.10	0.70	0.887	T348I	12.06	5.93×10^{-8}
3:50,356,435	rs140376758	0.000833	G	A	2	0.92	0.71	0.195	G321G	13.48	2.73×10^{-7}
3:50,357,035	[a]	0.000826	G	A	2	1.02	0.75	0.173	P296S	25.2	2.91×10^{-7}
3:50,357,036	rs140891320	0.000425	T	C	1	2.05	0.99	0.0388	R295R	0.021	9.74×10^{-7}
3:50,357,168	rs184202689	0.000414	A	G	1	1.46	1.00	0.144	S251S	0.004	3.39×10^{-7}
3:50,357,220	rs372371788	0.000828	C	T	2	2.17	0.74	0.00312	R234H	28.8	1.85×10^{-6}
3:50,357,223	[b]	0.000414	G	T	1	0.86	1.00	0.388	A233D	4.169	1.11×10^{-7}
3:50,357,244	rs374137877	0.000414	C	T	1	0.77	1.00	0.438	R226H	16.43	9.32×10^{-8}

[a] Not observed in dbSNP, Combined Annotation Dependent Depletion (CADD), or Exome Aggregation Consortium (ExAC)(44)

[b] Observed in ExAC(44)

[c] Gene-based p-value as calculated by Madsen-Browning (MB) test if variant is excluded

Abbreviations: Chr:pos, chromosome:position (GRCh37); MAF, minor allele frequency; A1, allele 1; A2, allele 2; SE, standard error; CADD, combined annotation dependent depletion score; MaxADP3, Maximal percent aggregation to 3.0 μM ADP

Table 4:

Association results of single variants in GSTZ1 used in gene-based tests for MaxEpi3 (n=521)

Chr:pos	rsID	MAF	A1	A2	N Carriers	Beta	SE	P-value	Mutation	CADD	Excl. p[b]
14:77,791,236	rs191650104	0.00096	A	C	1	3.04	0.99	0.00207	R13R	4.737	2.86×10^{-4}
14:77,793,206	[a]	0.00096	C	T	1	2.47	0.99	0.01284	Y31Y	10.79	5.02×10^{-5}
14:77,795,523	rs149972480	0.00096	G	A	1	2.58	0.99	0.00920	A134T	22.3	7.19×10^{-5}

[a] Observed in Exome Aggregation Consortium (ExAC)(44)

[b] Gene-based p-value as calculated by Madsen-Browning (MB) test if variant is excluded

Abbreviations: Chr:pos, chromosome:position (GRCh37); MAF, minor allele frequency; A1, allele 1; A2, allele 2; SE, standard error; CADD, combined annotation dependent depletion score; MaxEpi3, Maximal percent aggregation to 3.0 μM epinephrine