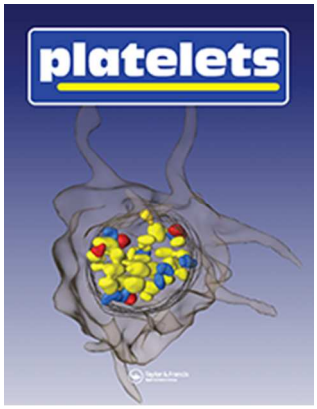


Exome-Chip Meta-Analysis Identifies Association between Variation in ANKRD26 and Platelet Aggregation

Journal:	*Platelets*
Manuscript ID	CPLA-2017-0197.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Chen, Ming-Huei; The Framingham Heart Study; National Heart, Lung, and Blood Institute, Division of Intramural Research, Population Sciences Branch Yanek, Lisa; Johns Hopkins University School of Medicine, GeneSTAR Research Program, Department of Medicine, Division of General Internal Medicine Backman, Joshua; University of Maryland School of Medicine, School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine Eicher, John; The Framingham Heart Study; National Heart, Lung, and Blood Institute, Division of Intramural Research, Population Sciences Branch Huffman, Jennifer; The Framingham Heart Study; National Heart, Lung, and Blood Institute, Division of Intramural Research, Population Sciences Branch Ben-Shlomo, Yoav; University of Bristol, School of Social and Community Medicine Beswick, Andrew; University of Bristol, School of Clinical Sciences Yerges-Armstrong, Laura; University of Maryland School of Medicine, School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine Schuldiner, Alan; University of Maryland School of Medicine, School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine O'Connell, Jeffrey; University of Maryland School of Medicine, School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine Mathias, Rasika; Johns Hopkins University School of Medicine, GeneSTAR

	Research Program, Department of Medicine, Divisions of Allergy and Clinical Immunology and General Internal Medicine Becker, Diane; Johns Hopkins University School of Medicine, GeneSTAR Research Program, Department of Medicine, Division of General Internal Medicine Becker, Lewis; Johns Hopkins University School of Medicine, GeneSTAR Research Program, Department of Medicine, Divisions of Cardiology and General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA Lewis, Joshua; University of Maryland School of Medicine, School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine Johnson, Andrew; The Framingham Heart Study; National Heart, Lung, and Blood Institute, Division of Intramural Research, Population Sciences Branch Faraday, Nauder; Johns Hopkins University School of Medicine, GeneSTAR Research Program, Department of Anesthesiology & Critical Care Medicine
Keywords:	platelets, platelet aggregation, SNP, genetic association, exome, platelet reactivity

Exome Chip Platelet Agg

Exome-Chip Meta-Analysis Identifies Association between Variation in *ANKRD26* and Platelet Aggregation

Ming-Huei Chen[1]*, Lisa R. Yanek[2]*, Joshua D. Backman[3]*, John D. Eicher[1], Jennifer E. Huffman[1], Yoav Ben-Shlomo[4], Andrew D. Beswick[5], Laura M. Yerges-Armstrong[3], Alan R. Shuldiner[3], Jeffrey R. O'Connell[3], Rasika A. Mathias[6], Diane M. Becker[2], Lewis C. Becker[7], Joshua P. Lewis[3]**, Andrew D. Johnson[1]**†, Nauder Faraday[8]**

*equal contribution
**equal supervision
†corresponding author

Correspondence: Dr. Andrew D. Johnson, 73 Mt Wayte Ave., Suite #2, Framingham, MA 01746, johnsonad2@nhlbi.nih.gov

[1]National Heart, Lung and Blood Institute's The Framingham Heart Study, Population Sciences Branch, Division of Intramural Research, National Heart, Lung and Blood Institute, Framingham, MA 01702 USA

[2]GeneSTAR Research Program, Department of Medicine, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[3]School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA

[4]School of Social and Community Medicine, University of Bristol, Bristol, UK

[5]School of Clinical Sciences, University of Bristol, Bristol, UK

[6]GeneSTAR Research Program, Department of Medicine, Divisions of Allergy and Clinical Immunology and General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[7]GeneSTAR Research Program, Department of Medicine, Divisions of Cardiology and General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[8]GeneSTAR Research Program, Department of Anesthesiology & Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Exome Chip Platelet Agg

Abstract

Previous genome-wide association studies (GWAS) have identified several variants associated with platelet function phenotypes; however, the proportion of variance explained by the identified variants is mostly small. Rare coding variants, particularly those with high potential for impact on protein structure/function, may have substantial impact on phenotype but are difficult to detect by GWAS. The main purpose of this study was to identify low frequency or rare variants associated with platelet function using genotype data from the Illumina HumanExome Bead Chip. Three family-based cohorts of European ancestry, including ~4,000 total subjects, comprised the discovery cohort, and two independent cohorts, one of European and one of African American ancestry, were used for replication. Optical aggregometry in platelet-rich plasma was performed in all the discovery cohorts in response to adenosine diphosphate (ADP), epinephrine, and collagen. Meta-analyses were performed using both gene-based and single nucleotide variant association methods. The gene-based meta-analysis identified a significant association ($P=7.13 \times 10^{-7}$) between rare genetic variants in *ANKRD26* and ADP-induced platelet aggregation. One of the *ANKRD26* SNVs - rs191015656, encoding a threonine to isoleucine substitution predicted to alter protein structure/function, was replicated in Europeans. Aggregation increases of ~20-50% were observed in heterozygotes in all cohorts. Novel genetic signals in *ABCG1* and *HCP5* were also associated with platelet aggregation to ADP in meta-analyses, although only results for *HCP5* could be replicated. The SNV in *HCP5* intersects epigenetic signatures in CD41+ megakaryocytes suggesting a new functional role in platelet biology for HCP5. This is the first study to use gene-based association methods from SNV array genotypes to identify rare variants related to platelet function. The molecular mechanisms and pathophysiological relevance for the identified genetic associations requires further study.

Exome Chip Platelet Agg

Platelets are blood cells that are critical to hemostasis after vascular injury, and their activation on ruptured atherosclerotic plaque contributes to pathologic thromboses in myocardial infarction and stroke[1]. We previously reported that platelet activation *in vitro*, in the presence and absence of antiplatelet therapy, is highly heritable[2, 3]. A number of subsequent genome-wide association studies (GWAS) identified several common genetic variants, including those in *PEAR1*, *GP6*, *ADRA2A*, and *JMJD1C*, which were related to platelet function[4, 5]; however, the overall proportion of phenotypic variance explained by the identified loci is modest except for GP6 in Jones et al[6]. Furthermore, most variants associated with platelet function, thus far, are noncoding, which has made it challenging to determine the mechanism(s) by which genotype influences phenotype.

Although GWAS have been successful in identifying associations between common genetic variants and platelet function phenotypes[4, 7-9], they are not well-suited to uncover low frequency (~1-5% minor allele frequency (MAF)) and rare variants (<1% MAF) that may have larger impact on phenotype. These less common variants are of interest when present in coding regions because they have greater potential to modify protein function, providing a clearer biological link between genetic and phenotypic variation. Indeed, by limiting genetic association analyses to exonic variants predicted to alter the gene product, Huffman et al[10] identified several low-frequency genetic variants in pathways not previously suspected of involvement in hemostasis. In another recent study, exome sequencing was used to discover exome-wide significant associations in *HYAL2* and *GSTZ1* with platelet aggregation in the Framingham Heart Study (FHS), with suggestive findings for *AR* and *MAPRE1*.[11]

The main objective of this study was to identify uncommon and rare genetic variants in coding sequences that are associated with platelet function. To achieve this aim we performed a meta-analysis of Exome Chip results from 3 family-based cohorts of European ancestry, in which platelet aggregation in response to several agonists was well-characterized. We used gene-based and single nucleotide variant (SNV) analyses to discover loci not previously known to be related to platelet function. Significant associations identified in our discovery meta-analysis were then replicated in independent cohorts of European or African American ancestry.

Methods

Description of study populations

GeneSTAR: The Genetic Study of Atherosclerosis Risk (GeneSTAR) is an ongoing, prospective study designed to determine environmental, phenotypic, and genetic causes of premature cardiovascular disease. Participants (n=3,003) were recruited from European- and African-American families (n=891) 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 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[2]. All measures described here were obtained prior to the commencement of aspirin therapy. Exclusion criteria included: 1) the presence of any coronary heart disease or vascular thrombotic event, 2) any bleeding disorder or any hemorrhagic event in the past (stroke, gastrointestinal bleed), 3) use of any anticoagulants or anti-platelet agents (i.e. warfarin, persantin, clopidogrel), 4) chronic 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 treatment trial, 8) recent menorrhagia, 9) known aspirin intolerance or allergic side effects, 10) serious medical disorders, including autoimmune diseases, renal or hepatic failure, cancer or HIV-AIDS, 11) chronic or acute use of glucocorticosteroid therapy or any drug that may interfere with the measured outcomes, 12) serious psychiatric disorders, and, 13) unable to independently make a decision to participate.

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 5,209 participants, the Offspring cohort (second generation), began to recruit in 1971, contains 5,124 participants, and the Third Generation cohort, began to recruit in 2002, contains 4,095 participants. In the present study we use data from the Offspring cohort[4] only.

PAPI Study: The Pharmacogenomics of Anti-Platelet Intervention (PAPI) study is a prospective cohort trial examining the relation of genetic variants to platelet function at baseline and in response to clopidogrel and aspirin. Characteristics of the Old Order Amish European cohort that comprise the PAPI study sample have been described previously[12]. 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.

Caerphilly Prospective Study (CaPS): The CaPS study is a sample of unrelated men of European ancestry recruited around the town of Caerphilly, South Wales as described in the past[9]. The ADP phenotype used in this study was collected during phase 2 of CaPS.

A fasting blood sample was drawn from all participants in each study. A complete blood count and cholesterol levels were determined and low-density lipoprotein cholesterol levels was calculated using the Friedewald equation. Hypertension was defined as systolic blood pressure ≥ 140 mm Hg, diastolic blood pressure ≥ 90 mm Hg, and/or use of prescription blood pressure lowering medications. Diabetes and current smoking status were obtained by self-report.

Each of these studies was approved by their local review board (GeneSTAR- Johns Hopkins Institutional Review Board; FHS- Boston University Institutional Review Board; and PAPI- University of Maryland Institutional Review Board) and written informed consent was obtained from all participants. Approval for the CaPS was granted by the South East Wales Local Research Ethics Committee Panel B.

Platelet function tests and phenotype harmonization:
Methods to assess platelet function *in vitro* are described in detail elsewhere[4, 12]. All 3 discovery cohorts (FHS, GS, PAPI) drew venous blood from fasting participants into citrated (3.2% or 3.8% in FHS) vacutainer tubes. In PAPI, the first 5mL of blood drawn were discarded. In GS, the first 4mL of blood drawn were discarded. All cohorts centrifuged to isolate PRP (180 x g for 15 min in GS and PAPI, 160 x g for 5 min in FHS). After transferring the PRP additional centrifugation was used to isolate PPP (2000 x g for 10 min in GS and PAPI, 2500 x g for 20 min

in FHS). All cohorts used no brake on centrifugation steps. In GS and PAPI, platelet counts were adjusted to 200,000 platelets/ul with PPP if the count was less than 200,000 platelets/ul. Platelet aggregation was assessed by traditional Born light transmission aggregometry using either a PAP-4 aggregometer (GS and FHS) or a PAP-8E aggregometer (PAPI) (Bio/Data Corporation, Horsham, Pennsylvania) at 37C with siliconized stir bars rotating at 1200 rpm.

Both GS and PAPI sourced their agonists from the Chronolog Corporation (Havertown, Pennsylvania) while FHS used collagen agonist from Bio/Data Corporation (Horsham, Pennsylvania) and ADP and epinephrine form Sigma Aldrich Corporation (St. Louis, Missouri). In GS, platelets were stimulated with ADP (2, 5, 10, and 20 μM), equine tendon-derived Type I collagen (1, 2, 5, and 10 μg/ml), epinephrine (2 and 10 μM), or arachidonic acid (1.6 mM). In PAPI, platelets were stimulated with the same agonists and concentrations except that 2 μM epinephrine was not measured. In FHS, the percent aggregation was determined after adding varying concentrations of ADP (0.05, 0.1, 0.5, 1.0, 3.0, 5.0, 10.0 and 15.0 μM) and epinephrine (0.01, 0.03, 0.05, 0.1, 0.5, 1.0, 3.0, 5.0 and 10.0 μM), and with a fixed concentration of arachidonic acid (5 mg/mL). In GS and PAPI, maximum aggregation within 5 min of agonist addition was measured as the percent change in light transmission relative to the PPP control from the same sample. In FHS, maximum percent aggregation was similarly calculated after 4 min for ADP and 5 min for epinephrine. Not all FHS samples were measured at all doses for ADP and epinephrine. Depending on responses to entry dose responses for ADP and epinephrine, additional higher or lower doses were added to reach at least a half-maximal response, or a low dose without a half-maximal response, resulting in different sample sizes at different doses. The lowest dose to create a 50% maximal aggregation (EC_{50}) to ADP or epinephrine was determined for all samples in this fashion. Collagen lag time to 1.9 μmol/L of calf-skin derived Type I collagen was also determined in FHS. As previously noted the distributions of collagen lag times differed somewhat between FHS and GS[4] which may be a factor of differing collagen sources and doses. Platelet aggregation to 0.725 uM ADP (Sigma Aldrich Corp, St. Louis, MO) was determined in CaPS using a Rubel-Renaud coaguloaggregometer at 37C with stirring at 1100 rpm. In CaPS 30 mL of blood was drawn for other uses before fasting blood was drawn into 0.13M sodium citrate vacutainers for the platelet studies. After isolation, PRP was diluted to 300,000 platelets/*u*l in autologous PPP. Primary and secondary (maximal after ~2-2.5 min) percent aggregation were measured.

Phenotypes were harmonized across studies such that twelve traits were defined for analysis: two low-dose ADP traits, two high-dose ADP traits, two low-dose epinephrine traits, two high-dose epinephrine traits, and four collagen traits. Details are shown in **Supplemental Table 1**.

Exome Chip, genotype calls, and quality control (Q/C):
Samples were genotyped using the Illumina HumanExome Bead Chip (GeneSTAR and PAPI used v1.2; FHS used v1.0) and cluster files were created using GenomeStudio software. Genotypes were called using a similar procedure as reported by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Exome Chip effort[13]. The total number of variants for analysis was 246,670, including monomorphic variants from autosomes and X chromosome. Quality control procedures were similar to those reported in an Exome Chip meta-analysis of the Blood Cell Consortium[14]. All variants in each study were aligned to the forward strand and an indexed variant call file for subsequent association analyses was generated

and checked for allele alignment using the checkVCF package (https://github.com/zhanxw/checkVCF). Cohort level association results were QC'd prior to meta-analysis using EasyQC software (www.genepi-regensburg.de/easyqc)[15].

Genotyping and imputation in the Caerphilly Prospective study

Genotyping in CaPS was done with the Affymetrix UK BioBank array with genotype calling using the Affymetrix Axiom Analysis Suite software. After sample quality control procedures (excluding duplicate/reference samples, samples with heterozygosity rate >6 standard deviation from the mean and possible sex mismatches) and genotyping quality control procedure (excluding variants with Hardy Weinberg Equilibrium P<1E-6, minor allele frequency < 1%), the final dataset for genotype imputation consisted of 1,184 samples and 646,137 variants. The imputation of 22 autosomes was done using the Haplotype Reference Consortium release 1.1, 2016 reference panel (European ancestry) with the University of Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html).[16]

Statistical analysis

Inverse normal transformation was applied to all platelet function traits in each cohort. Traits were adjusted for age, sex, and population stratification-based principal components. In each cohort, the associations of autosomal and X-chromosomal variants with platelet phenotype were analyzed by either raremetalworker (http://genome.sph.umich.edu/wiki/RAREMETALWORKER) or rvtests (http://zhanxw.github.io/rvtests/)[17] with familial correlation accounted for using a linear mixed effects model. The single variant summary statistics and covariance matrices generated from raremetalworker and rvtests were used for meta-analysis. All meta-analyses were performed separately by two independent analysts (MHC and LRY) and the results obtained were identical.

The discovery cohort consisted of Europeans in the GeneSTAR, FHS, and PAPI studies. Meta-analyses were used to discover variants associated with the harmonized platelet phenotypes using both gene-based and SNV approaches. Meta-analyses were performed using RAREMETAL software (http://genome.sph.umich.edu/wiki/RAREMETAL)[18]. Variants were included in meta-analyses if the genotype call rate was ≥95%, Hardy-Weinberg equilibrium p-values >$1x10^{-7}$, and allele frequency among cohorts differed by <0.3. SNV statistics were meta-analyzed using the Cochran-Mantel-Haenszel method. Variable threshold (VT) approach and sequence kernel association tests (SKAT) implemented in RAREMETAL were used for gene-based tests. VT has better power when selected rare variants in a gene have the same effect direction[19], while SKAT has better power when selected variants have different effect directions[20]. For gene-based tests, we considered only missense, nonsense, and splice site variants with MAF < 0.01. Bonferroni correction was applied to determine genome-wide significant thresholds for gene-based and SNV meta-analyses, where the total number of tests was defined by the number of analyzed genes/variants for each platelet reactivity trait. The thresholds were ~$3.3x10^{-6}$ for gene-based and ~$4.3x10^{-7}$ for SNV meta-analyses. When a genome-wide significance signal was identified, conditional analyses were conducted (at the meta-analysis level) using RAREMETAL to identify independent signals. The selected significant variants were tested for independence in each round and the procedure was repeated until there was no new signal.

Exome Chip Platelet Agg

Two independent cohorts were used to replicate variants identified from the discovery meta-analyses: 1) CaPS, an all-male population sample of European ancestry (for ADP 0.725 *u*M in PRP phenotype) [9], and, 2) the GeneSTAR African American cohort (for all phenotypes). For CaPS, Haplotype Reference Consortium imputed data (based on HRCr1.1 2016 reference panel) were used for replication. SNVs from genes that met Bonferroni significance in the gene-based meta-analysis were selected for replication if the individual SNVs also met nominal significance and the variant was predicted by PolyPhen score[21] and/or SIFT to damage protein structure/function. Consistent with previous methods used to replicate rare variants from locus-based discovery methods, these rare exonic SNVs were selected because of the high probability that these variants had a direct impact on protein function[22, 23].

Results

Sample characteristics of the discovery and replication cohorts including coronary heart disease risk factors are shown on **Table 1**. In general, the study cohorts consisted of middle-aged adults, approximately half of whom were women (except CaPS which is solely male).

QQ plots with genomic control estimates from cohort level GWAS, SNV meta-analyses and gene-based meta-analyses can be found in **Supplemental Figure 1**, where the results show no systematic inflation in our meta-analyses.

Gene-based meta-analyses and relation of rare nonsynonymous coding variants to platelet function

Gene-based meta-analyses were used to discover rare exonic variants, in up to 15,478 genes associated with platelet function phenotype. **Supplemental Table 2** shows top results from the gene-based association analysis. Only 2 genes (*ANKRD26, ABCG1*) met Bonferroni-corrected significance for association with platelet function.

Genetic variation in *ANKRD26* was significantly associated with low dose ADP-induced platelet aggregation using VT gene-based analyses (**Table 2**). Twelve rare SNVs (cumulative MAF=0.012) in *ANKRD26*, none of which were present in the PAPI cohort, contributed to the association (β= 0.619; P=7.13x10^{-7}). The MAFs for the 12 SNVs was <0.3% in our European sample, and occurrence of the minor alleles of the 12 SNVs appeared sporadically rather than clustered within families (**Table 2**). The individual SNVs in *ANKRD26* were independent of each other, with maximum pairwise linkage disequilibrium (r^2) among them of < 6.58x10^{-6} (**Supplemental Table 3**), and together they contributed to an overall positive association with ADP aggregation. For each SNV present in both cohorts, the direction of effect was the same in GeneSTAR and FHS. Five of the variants met nominal significance for association with low dose ADP-induced aggregation (**Table 2**); however, no single variant met Bonferroni threshold for genome-wide significance in SNV analyses.

We sought to replicate gene-based findings for specific SNVs identified in *ANKRD26* gene. Three of the 5 significant SNVs in *ANKRD26* were predicted to be possibly or probably damaging by PolyPhen score (**Table 3**). Four of these 5 SNVs were predicted deleterious by

Exome Chip Platelet Agg

SIFT, including the 3 predicted possibly or probably damaging by PolyPhen. Of these 5 SNVs, rs191015656 was significantly associated with platelet aggregation to 0.725 uM ADP in CaPS, and the direction of effect was the same as that observed in GeneSTAR and FHS. Carriers of one copy of the minor allele (A) at rs191015656 had ~17%, ~33%, and ~50% greater aggregation to ADP 3 uM, ADP 2 uM, and ADP 0.725 uM phenotype (in FHS, GeneSTAR, CaPS, respectively) than GG homozygotes (**Fig. 1**). This SNV was not present in the GeneSTAR AA or PAPI cohorts. The SNV causes a Thr181Ile change in the protein and overlaps the ANK5 protein domain according to UniProt annotation. The other SNVs in *ANKRD26* did not replicate or were not present in CaPS or GeneSTAR African Americans (**Table 3**).

Three SNVs in *ABCG1*, each occurring in <0.03% of our sample, contributed to a statistically significant gene-based association with high dose ADP-induced platelet aggregation (β= -2.831; $P=1.92x10^{-06}$). Pairwise r^2 indicated no linkage disequilibrium among the SNVs. Overall, the variants were associated with reduced ADP aggregation. Two variants met nominal significance for association with ADP aggregation (**Table 2**), but none met genome-wide significant threshold in SNV analyses. We attempted to replicate findings from the 2 SNVs in *ABCG1* that met nominal significance; however, neither SNV could be imputed in CaPS with high confidence (for rs151254598 imputation RSqr=0.00683) nor were they present in our GeneSTAR AA cohort (**Table 3**).

SNV meta-analyses and relation of common variants to platelet function

We also meta-analyzed the SNVs present on the Exome Chip to uncover novel variants associated with platelet function. After meta-analysis of variants present in 2 or more of the discovery samples a total of up to 129,094 SNVs were examined. **Supplemental Table 4** shows association results for the top 55 SNVs with $P<7x10^{-06}$, excluding the *HCP5* and *TRIM24* findings discussed below.

The SNV meta-analysis identified 11 variants in 7 chromosomal regions that were genome-wide significantly associated with platelet function phenotypes. All SNVs identified were common- the MAF for each was >9% in our meta-analysis cohort of European ancestry. Association findings for 5 of the SNVs- rs12566888 in *PEAR1*, rs1671152 in *GP6*, rs869244 and rs4545476 near *ADRA2A*, rs342293 near *PIK3CG*, and rs10761731 in *JMJD1C*- confirm our previous reports and are not considered further here (**Supplemental Table 5**)[4]. Three additional SNVs in *GP6* were found to be associated with platelet response to collagen (**Supplemental Table 4**) and all were nonsynonymous. However, these SNVs were all in LD ($r^2>0.76$) with another SNV in GP6- rs1671152- that we previously reported to be associated with platelet response to collagen[4].

Several of the SNV association findings with platelet activation phenotypes were novel. rs2263316, which is an intronic variant in a long non-coding RNA called HLA complex P5 (HCP5), was associated with platelet aggregation to ADP (**Table 4**; $P= 3.45x10^{-07}$). The alternate allele was related to greater platelet aggregation in the 3 European cohorts included in the meta-analysis. In the CaPS and GeneSTAR African American replication cohorts, the direction of effect for this SNV was the same achieving nominal significance for replication (P=0.044 and 0.039 for CaPS and GeneSTAR African Americans, respectively; **Table 4**). Employing BLUEPRINT Consortium regulatory annotations, we found that rs2263316 intersects with

CD34-CD41+CD42+ megakaryocytic cell peaks for hypermethylation and H3K9me3 histone marks, and nearly adjacent to an H3K27me3 peak. The same region is also a DNase hotspot in CD14+CD16- monocytes and overlaps H3K27me3 marks in mesenchymal stem cell bone marrow.

Novel associations with epinephrine-induced platelet aggregation were identified for an intronic variant of *TRIM24* (rs1874326) (**Table 4**). The alternate allele at this locus was associated with reduced platelet aggregation to epinephrine in all 3 European cohorts included in the meta-analyses. However, the finding in European cohorts could not be replicated in GeneSTAR African-Americans individuals and epinephrine-induced platelet aggregation was not assessed in CaPS. The SNV overlaps CD34-CD41+CD42+ megakaryocytic peaks for hypermethylation, H3K27ac and H3K4me1 histones from the BLUEPRINT Consortium.

Discussion

The main finding of this Exome Chip meta-analysis was discovery and replication of an association between a rare coding SNV in *ANKRD26* and ADP-induced platelet aggregation. Additional associations of SNVs in *ABCG1* and *HCP5* with platelet aggregation were also suggested in meta-analyses, with *HCP5* findings nominally confirmed in two additional population samples.

We previously performed GWAS to identify several common genetic variants associated with agonist-induced platelet aggregation[4, 5]. The SNV meta-analysis in this study confirmed these previous reports, including the following genetic association findings: *PEAR1* with ADP- and epinephrine-induced platelet aggregation; *JMJD1C* with epinephrine-induced platelet aggregation; *PIK3CG* with epinephrine-induced platelet aggregation; *ADRA2A* with epinephrine-induced platelet aggregation; and *GP6* with collagen-induced platelet aggregation.

To our knowledge, this is the first report describing an Exome Chip-based study to identify the genetic basis for variability in platelet function. Advantages of an Exome Chip approach over GWAS are greater potential to identify genotype-phenotype associations for rare variants and a closer biological connection between genetic variants and alterations in protein structure/function. The coverage with Exome Chip design provides a different view of the exome than sequencing-based approaches (e.g., with respect to platelet aggregation[11]). While the Exome Chip has the limitation that it does not find and call *de novo* variants, exome sequencing may miss variants present on the Exome Chip due to limitations in capture reagent sets and variable site coverage. To enhance our ability to detect biologically relevant rare genetic associations, we limited our gene-based analyses to missense, nonsense, and splice site variants with MAF < 0.01. A similar approach was used to identify novel variants contributing to platelet count and mean platelet volume[14]. Furthermore, we considered the impact of the variant on protein structure/function using PolyPhen[21] and SIFT scores when selecting specific variants for replication.

We identified a strong relation between genetic variation in *ANKRD26* and platelet aggregation to ADP in our gene-based meta-analysis. Several rare variants contributed to this overall gene-based effect; among these, 5 SNVs were nominally significant in the discovery meta-analysis

and 3 of these 5 were predicted to alter protein structure/function with two prediction programs. One of these SNVs- rs191015656- is a single nucleotide substitution of adenine for guanine, causing isoleucine to be substituted for threonine in the primary protein structure. The Thr181Ile variant we discovered to be associated with increased platelet aggregation is within the fifth Ankyrin repeat domain. Ankyrin repeats are often sites of protein-protein interactions. This specific variant was significantly associated with platelet aggregation to ADP in the CaPS replication cohort, and greater aggregation to ADP was observed for its heterozygotes in all cohorts studied. The magnitude of effect of this genetic variant on aggregation appeared greatest- 50% higher compared to GG homozygotes- for the cohort treated with the lowest dose of ADP. Allelic effects on platelet aggregation may be more likely to be unmasked at low doses of agonist where the external stimulus does not overwhelm the endogenous gene effects. This may suggest that future studies seeking to map genetic determinants of platelet aggregation in the general population may want to include assessment of low agonist doses.

The mechanism underlying an association between genetic variation in *ANKRD26* and enhanced platelet aggregation to ADP cannot be determined from this study. However, *ANKRD26* has previously been reported to be involved in thrombopoiesis. *ANKRD26* encodes a 192kD protein that is highly expressed in megakaryocytes (MKs)[24], and several mutations in the 5' untranslated regulatory region (UTR) are associated with a rare form of familial autosomal dominant thrombocytopenia[25]. Functional studies indicate transcription factors (RUNX1/FLI1) fail to bind the mutated 5' UTR of *ANKRD26* and suppress its expression, and its persistence during MK development leads to a reduction in proplatelet formation[26]. It was recently reported that some familial mutations in the 5' UTR of *ANKRD26* in subjects with hereditary thrombocytopenia are also associated with reduced platelet aggregation to arachidonic acid and epinephrine[27]. However, a study of 22 independent individuals with 5'UTR mutations found no specific platelet aggregation defect[287]. A recent Exome Chip meta-analysis of >150,000 subjects did not reveal a relation between coding variants in *ANKRD26* and platelet count or mean platelet volume[14] and an association with platelet aggregation was not investigated. The role of upstream mutations and coding mutations may be distinct in their effects on platelet maturation, thrombocytopenia and platelet function. Additional studies are needed to elucidate the molecular mechanism for the relation between exonic variation in *ANKRD26* and platelet aggregation and the clinical consequences, if any, of this variation.

ABCG1 encodes an ATP-binding cassette (ABC) transporter known to be involved in cellular efflux of high density lipoprotein, suppression of toll-like receptor mediated inflammation, and atherosclerosis[29]. ABC transporters are also known to regulate maturation of hematopoietic progenitor cells and MKs[29]; however, an association with platelet aggregation has not been reported. Although our gene-based meta-analysis suggested an association between genetic variation in *ABCG1* and ADP-induced aggregation, PolyPhen and SIFT predicted both variants we identified to be benign. One of these variants- rs151254598- encoding a glycine to arginine substitution, was associated with a reduction in ADP-induced aggregation in CaPS similar in direction to the discovery cohort; however, we could not impute this rare SNV in CaPS with confidence. Thus, confirmation of this genotype-phenotype association requires additional replication.

Exome Chip Platelet Agg

Our SNV meta-analysis identified a common intronic variant in *HCP5* associated with ADP-induced aggregation. We observed modest replication of this finding in CaPS and the GeneSTAR AA cohorts (P=0.044 and 0.039, respectively). *HCP5* encodes a long non-coding RNA within a region that is rich in DNAse I hypersensitivity sites, which are particularly notable in cells of hematopoietic origin. Using BLUEPRINT Consortium data, we show direct overlap of rs2263316 with CD34-CD41+CD42+ megakaryocytic cell peaks for hypermethylation and H3K9me3 histone marks. This SNV is also nearly adjacent to an H3K27me3 peak. These histone marks generally reflect repressive regulatory effects and suggest HCP5 may have specific roles and expression in megakaryocytes. From public gene expression resources, tissues with the highest RNA expression of *HCP5* include bone marrow, whole blood, immune cells and spleen. Interestingly, SNVs in *HCP5* are associated with disease progression and viral load in patients with HIV infection, suggesting an immunomodulatory role for *HCP5*[30]; however, an association with platelet function has not previously been described. Additional studies are required to confirm this association and determine its molecular basis.

Our study has several limitations. We focused on exonic variants, particularly those affecting protein structure/function, to enhance biological plausibility for genotype-phenotype association findings; however, we did not study the molecular mechanism for findings and we cannot exclude that associations observed are due to LD with ungenotyped exonic, intronic, or intergenic variants. Our discovery cohort included ~4,000 subjects of European descent for which we had sufficient power to detect an effect for a rare variant (i.e., occurring with MAF<0.01) explaining 1% of trait variance with 81.3% power (**Supplemental Table 6**). Thus, we could have missed gene-association findings for variants occurring less frequently or with smaller effect size. We examined genetic associations with platelet function in response to ADP, epinephrine, and collagen. Although we harmonized platelet phenotypes, heterogeneity of phenotypic assessment across studies may have impeded our ability to detect important genotype-phenotype associations. Finally, although platelet aggregation phenotypes have been reported to be associated with thrombotic outcomes (e.g. myocardial infarction) in human subjects[31, 32], the pathophysiologic relevance of the genotype-phenotype associations identified in this report are yet unknown. Strengths of the study include that it is one of the largest genetic studies of platelet aggregation to date, and the inclusion of two replication samples of diverse origin.

In summary, in a gene-based meta-analysis involving ~4,000 individuals of European ancestry, we identified a novel association between a rare variant in *ANKRD26* (rs191015656) and greater platelet aggregation to ADP. A novel association in *HCP5* (rs2263316) with greater platelet aggregation to ADP was also identified. Supportive evidence for these associations was provided in independent replication samples. The molecular mechanism and clinical consequences for these gene association findings require further study.

Exome Chip Platelet Agg

Tables

Table 1: Sample characteristics of the study cohorts.

Table 2: Significant association findings from gene-based meta-analysis in discovery cohorts.

*Phenotype = ADP low 1; **Phenotype= ADP high 2; MAF= minor allele frequency; Ref= reference allele; Alt= alternate allele; Nhet= number of heterozygotes; Nhetfam= number of families with at least one heterozygote; direction of effect appears in order -GS/FHS/PAPI, X = variant not present in cohort; ***alternate protein isoforms for ANKRD26 and ABCG1 exist that affect some or all of the relative amino acid positions

Table 3: Translational effect and replication of gene-based association findings.

MAF= minor allele frequency; Rsq= imputation accuracy (observed: expected) measure; bold highlight denotes significant replication finding; NA = no information available (no allele carriers) *alternate protein isoforms for ANKRD26 and ABCG1 exist that affect some or all of the relative amino acid positions

Table 4: Significant novel locus findings and replication from single nucleotide variant analysis.

Bold highlight denotes significant replication finding; *see Supplemental Table 1 for meta-analysis phenotype descriptions and mean (SD) in each cohort; **1-sided test statistics are reported since replication required *a priori* the same direction of effect as discovery; Epinephrine data was not acquired in CaPS and thus unavailable for replication of the TRIM24 signal

Figure legends

Figure 1. Impact of genetic variation in *ANKRD26* (rs191015656) on ADP-induced platelet aggregation in discovery and replication cohorts. Optical aggregation was performed in platelet rich plasma in response to ADP. The concentrations of ADP used to induce aggregation (from left to right) were as follows: 0.725 uM, 3 uM, and 2 uM, in Caerphilly Prospective Study (CaPS), the Framingham Heart Study (FHS), and GeneSTAR (GS), respectively.

Supplemental Figure 1. QQ plots and genomic enrichment values for all single variant and gene-based tests in individual cohorts and meta-analyses.

Supplemental Tables

Supplemental Table 1. Harmonization of platelet phenotype for meta-analysis.

Supplemental Table 2. Gene-based analysis suggestive signals (P<1.0E-04).

Exome Chip Platelet Agg

Supplemental Table 3. Linkage disequilibrium calculations among *ANKRD26* variants in unrelated individuals in FHS.

Supplemental Table 4. SNV known or suggestive signals (P<7.0E-06).

Supplemental Table 5. SNV associations confirming prior reported genome-wide significant findings[4].

Supplemental Table 6. Power analysis for 4,000 individuals with family structure. We used SOLAR[33] to simulate phenotypes conditional on the observed family structures in the FHS sample. The phenotypes were simulated to have variance of 1 and polygenic heritability of 0.25. We considered the QTL (additive model) with MAF of 0.01, 0.05, 0.1, 0.2, 0.3 and 0.4, and QTL variance of 0.0025, 0.005, 0.0075 and 0.01. 1,000 of replicates were used to estimate power for each considered scenario. Simulated phenotypes were analyzed using a linear mixed effects model implemented in the *lmekin* function from the *coxme* R package where a relationship coefficient matrix was used to account for familial correlation.

Acknowledgments

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). Genotyping, quality control, and calling of the Illumina HumanExome BeadChip in the Framingham Heart Study were supported by funding from the National Heart, Lung and Blood Institute Division of Intramural Research (Daniel Levy and Christopher J. O'Donnell, Principal Investigators). Support for the centralized genotype calling was provided by Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium through the National Institutes of Health (NIH) American Recovery and Reinvestment Act of 2009 (5RC2HL102419). M.H.C., J.D.E. and A.D.J. were supported by National Heart, Lung and Blood Institute Division of Intramural Research funds.

GeneSTAR was supported by the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518, HL087698, and 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. Genotyping services were provided through the RS&G Service by the Northwest Genomics Center at the University of Washington, Department of Genome Sciences, under U.S. Federal Government contract number HHSN268201100037C from the National Heart, Lung, and Blood Institute.

The Caerphilly Prospective study was undertaken by the former MRC Epidemiology Unit (South

Exome Chip Platelet Agg

Wales) and was funded by the Medical Research Council of the United Kingdom. The data archive is maintained by the School of Social and Community Medicine, University of Bristol.

This investigation was also supported by National Institutes of Health grants U01 GM074518, U01 HL105198, K23 GM102678, and the University of Maryland Mid-Atlantic Nutrition and Obesity Research Center (P30 DK072488).

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.

Exome Chip Platelet Agg

References

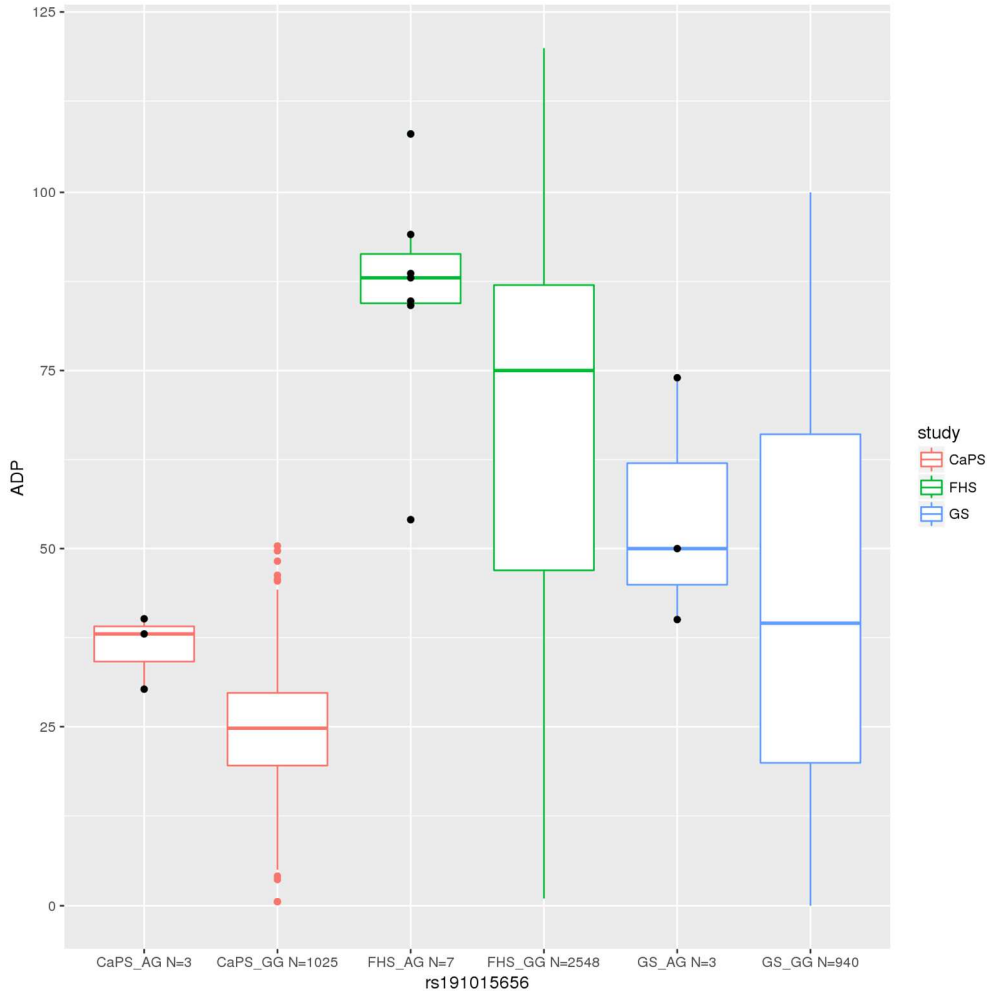
1. Patrono C, Rodriguez LAG, Landolfi R, Baigent C. Low-dose aspirin for the prevention of atherothrombosis. *New England Journal of Medicine*. 2005;353:2373-2383
2. Faraday N, Yanek LR, Mathias R, Herrera-Galeano JE, Vaidya D, Moy TF, Fallin MD, Wilson AF, Bray PF, Becker LC, Becker DM. Heritability of platelet responsiveness to aspirin in activation pathways directly and indirectly related to cyclooxygenase-1. *Circulation*. 2007;115:2490-2496
3. Bray PF, Mathias RA, Faraday N, Yanek LR, Fallin MD, Herrera-Galeano JE, Wilson AF, Becker LC, Becker DM. Heritability of platelet function in families with premature coronary artery disease. *J Thromb Haemost*. 2007;5:1617-1623
4. Johnson AD, Yanek LR, Chen MH, Faraday N, Larson MG, Tofler G, Lin SJ, Kraja AT, Province MA, Yang Q, Becker DM, O'Donnell CJ, Becker LC. Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. *Nature genetics*. 2010;42:608-613
5. Lewis JP, Ryan K, O'Connell JR, Horenstein RB, Damcott CM, Gibson Q, Pollin TI, Mitchell BD, Beitelshees AL, Pakzy R, Tanner K, Parsa A, Tantry US, Bliden KP, Post WS, Faraday N, Herzog W, Gong Y, Pepine CJ, Johnson JA, Gurbel PA, Shuldiner AR. Genetic variation in pear1 is associated with platelet aggregation and cardiovascular outcomes. *Circ Cardiovasc Genet*. 2013;6:184-192
6. Jones, C. I., Garner, S. F., Angenent, W., Bernard, A., Berzuini, C., Burns, P., Farndale, R. W., Hogwood, J., Rankin, A., Stephens, J. C., Tom, B. D., Walton, J., Dudbridge, F., Ouwehand, W. H., Goodall, A. H. and On behalf of the Bloodomics Consortium (2007), Mapping the platelet profile for functional genomic studies and demonstration of the effect size of the *GP6* locus. Journal of Thrombosis and Haemostasis, 5: 1756–1765. doi:10.1111/j.1538-7836.2007.02632.x
7. Mathias RA, Kim Y, Sung H, Yanek LR, Mantese VJ, Hererra-Galeano JE, Ruczinski I, Wilson AF, Faraday N, Becker LC, Becker DM. A combined genome-wide linkage and association approach to find susceptibility loci for platelet function phenotypes in ceuropean american and african american families with coronary artery disease. *BMC Med Genomics*. 2010;3:22
8. Qayyum R, Becker LC, Becker DM, Faraday N, Yanek LR, Leal SM, Shaw C, Mathias R, Suktitipat B, Bray PF. Genome-wide association study of platelet aggregation in african americans. *BMC Genet*. 2015;16:58
9. Eicher JD, Xue L, Ben-Shlomo Y, Beswick AD, Johnson AD. Replication and hematological characterization of human platelet reactivity genetic associations in men from the caerphilly prospective study (caps). *Journal of Thrombosis and Thrombolysis*. 2016;41:343-350
10. Huffman JE, de Vries PS, Morrison AC, Sabater-Lleal M, Kacprowski T, Auer PL, Brody JA, Chasman DI, Chen MH, Guo X, Lin LA, Marioni RE, Muller-Nurasyid M, Yanek LR, Pankratz N, Grove ML, de Maat MP, Cushman M, Wiggins KL, Qi L, Sennblad B, Harris SE, Polasek O, Riess H, Rivadeneira F, Rose LM, Goel A, Taylor KD, Teumer A, Uitterlinden AG, Vaidya D, Yao J, Tang W, Levy D, Waldenberger M, Becker DM, Folsom AR, Giulianini F, Greinacher A, Hofman A, Huang CC, Kooperberg C, Silveira A, Starr JM, Strauch K, Strawbridge RJ, Wright AF, McKnight B, Franco OH, Zakai N, Mathias RA, Psaty BM, Ridker PM, Tofler GH, Volker U, Watkins H, Fornage M, Hamsten A, Deary IJ, Boerwinkle E, Koenig W, Rotter JI, Hayward C,

Dehghan A, Reiner AP, O'Donnell CJ, Smith NL. Rare and low-frequency variants and their association with plasma levels of fibrinogen, fvii, fviii, and vwf. *Blood*. 2015;126:e19-29

11. Eicher JD, Chen MH, Pitsillides AN, Lin H, Veeraraghavan N, Brody JA, Metcalf GA, Muzny DM, Gibbs RA, Becker DM, Becker LC, Faraday N, Mathias RA, Yanek LR, Boerwinkle E, Cupples LA, Johnson AD. Whole exome sequencing in the framingham heart study identifies rare variation in hyal2 that influences platelet aggregation. *Thrombosis and haemostasis*. 2017
12. Shuldiner AR, O'Connell JR, Bliden KP, Gandhi A, Ryan K, Horenstein RB, Damcott CM, Pakyz R, Tantry US, Gibson Q, Pollin TI, Post W, Parsa A, Mitchell BD, Faraday N, Herzog W, Gurbel PA. Association of cytochrome p450 2c19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA*. 2009;302:849-857
13. Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, Hansen M, Borecki IB, Cupples LA, Fornage M, Gudnason V, Harris TB, Kathiresan S, Kraaij R, Launer LJ, Levy D, Liu Y, Mosley T, Peloso GM, Psaty BM, Rich SS, Rivadeneira F, Siscovick DS, Smith AV, Uitterlinden A, van Duijn CM, Wilson JG, O'Donnell CJ, Rotter JI, Boerwinkle E. Best practices and joint calling of the humanexome beadchip: The charge consortium. *PLoS One*. 2013;8:e68095
14. Eicher JD, Chami N, Kacprowski T, Nomura A, Chen MH, Yanek LR, Tajuddin SM, Schick UM, Slater AJ, Pankratz N, Polfus L, Schurmann C, Giri A, Brody JA, Lange LA, Manichaikul A, Hill WD, Pazoki R, Elliot P, Evangelou E, Tzoulaki I, Gao H, Vergnaud AC, Mathias RA, Becker DM, Becker LC, Burt A, Crosslin DR, Lyytikainen LP, Nikus K, Hernesniemi J, Kahonen M, Raitoharju E, Mononen N, Raitakari OT, Lehtimaki T, Cushman M, Zakai NA, Nickerson DA, Raffield LM, Quarells R, Willer CJ, Peloso GM, Abecasis GR, Liu DJ, Deloukas P, Samani NJ, Schunkert H, Erdmann J, Fornage M, Richard M, Tardif JC, Rioux JD, Dube MP, de Denus S, Lu Y, Bottinger EP, Loos RJ, Smith AV, Harris TB, Launer LJ, Gudnason V, Velez Edwards DR, Torstenson ES, Liu Y, Tracy RP, Rotter JI, Rich SS, Highland HM, Boerwinkle E, Li J, Lange E, Wilson JG, Mihailov E, Magi R, Hirschhorn J, Metspalu A, Esko T, Vacchi-Suzzi C, Nalls MA, Zonderman AB, Evans MK, Engstrom G, Orho-Melander M, Melander O, O'Donoghue ML, Waterworth DM, Wallentin L, White HD, Floyd JS, Bartz TM, Rice KM, Psaty BM, Starr JM, Liewald DC, Hayward C, Deary IJ, Greinacher A, Volker U, Thiele T, Volzke H, van Rooij FJ, Uitterlinden AG, Franco OH, Dehghan A, Edwards TL, Ganesh SK, Kathiresan S, Faraday N, Auer PL, Reiner AP, Lettre G, Johnson AD. Platelet-related variants identified by exomechip meta-analysis in 157,293 individuals. *Am J Hum Genet*. 2016;99:40-55
15. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R, Ferreira T, Fall T, Graff M, Justice AE, Luan J, Gustafsson S, Randall JC, Vedantam S, Workalemahu T, Kilpelainen TO, Scherag A, Esko T, Kutalik Z, Heid IM, Loos RJ. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9:1192-1212
16. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. Next-generation genotype imputation service and methods. *Nature genetics*. 2016;48:1284-1287

17. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. Rvtests: An efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics*. 2016;32:1423-1426
18. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, Nikpay M, Auer PL, Goel A, Zhang H, Peters U, Farrall M, Orho-Melander M, Kooperberg C, McPherson R, Watkins H, Willer CJ, Hveem K, Melander O, Kathiresan S, Abecasis GR. Meta-analysis of gene-level tests for rare variant association. *Nature genetics*. 2014;46:200-204
19. Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet*. 2010;86:832-838
20. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011;89:82-93
21. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using polyphen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7 20
22. Liu DJ, Leal SM. Replication strategies for rare variant complex trait association studies via next-generation sequencing. *Am J Hum Genet*. 2010;87:790-801
23. Panoutsopoulou K, Tachmazidou I, Zeggini E. In search of low-frequency and rare variants affecting complex traits. *Human Molecular Genetics*. 2013;22:R16-21
24. Macaulay IC, Tijssen MR, Thijssen-Timmer DC, Gusnanto A, Steward M, Burns P, Langford CF, Ellis PD, Dudbridge F, Zwaginga JJ, Watkins NA, van der Schoot CE, Ouwehand WH. Comparative gene expression profiling of in vitro differentiated megakaryocytes and erythroblasts identifies novel activatory and inhibitory platelet membrane proteins. *Blood*. 2007;109:3260-3269
25. Pippucci T, Savoia A, Perrotta S, Pujol-Moix N, Noris P, Castegnaro G, Pecci A, Gnan C, Punzo F, Marconi C, Gherardi S, Loffredo G, De Rocco D, Scianguetta S, Barozzi S, Magini P, Bozzi V, Dezzani L, Di Stazio M, Ferraro M, Perini G, Seri M, Balduini CL. Mutations in the 5' utr of ankrd26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, thc2. *Am J Hum Genet*. 2011;88:115-120
26. Bluteau D, Balduini A, Balayn N, Currao M, Nurden P, Deswarte C, Leverger G, Noris P, Perrotta S, Solary E, Vainchenker W, Debili N, Favier R, Raslova H. Thrombocytopenia-associated mutations in the ankrd26 regulatory region induce mapk hyperactivation. *Journal of Clinical Investigation*. 2014;124:580-591
27. Perez Botero J, Chen D, He R, Viswanatha DS, Majerus JA, Coon LM, Nguyen PL, Reichard KK, Oliveira JL, Tefferi A, Gangat N, Pruthi RK, Patnaik MM. Clinical and laboratory characteristics in congenital ankrd26 mutation-associated thrombocytopenia: A detailed phenotypic study of a family. *Platelets*. 2016;27:712-715
28. Noris P, Perrotta S, Seri M, Pecci A, Gnan C, Loffredo G, Pujol-Moix N, Zecca M, Scognamiglio F, De Rocco D, Punzo F, Melazzini F, Scianguetta S, Casale M, Marconi C, Pippucci T, Amendola G, Notarangelo LD, Klersy C, Civaschi E, Balduini CL, Savoia A. Mutations in ankrd26 are responsible for a frequent form of inherited thrombocytopenia: Analysis of 78 patients from 21 families. *Blood*. 2011;117:6673-6680
29. Westerterp M, Bochem AE, Yvan-Charvet L, Murphy AJ, Wang N, Tall AR. Atp-binding cassette transporters, atherosclerosis, and inflammation. *Circulation Research*. 2014;114:157-170

30. Catano G, Kulkarni H, He W, Marconi VC, Agan BK, Landrum M, Anderson S, Delmar J, Telles V, Song L, Castiblanco J, Clark RA, Dolan MJ, Ahuja SK. Hiv-1 disease-influencing effects associated with znrd1, hcp5 and hla-c alleles are attributable mainly to either hla-a10 or hla-b*57 alleles. *PLoS One*. 2008;3:e3636
31. Breet NJ, van Werkum JW, Bouman HJ, Kelder JC, Ruven HJ, Bal ET, Deneer VH, Harmsze AM, van der Heyden JA, Rensing BJ, Suttorp MJ, Hackeng CM, ten Berg JM. Comparison of platelet function tests in predicting clinical outcome in patients undergoing coronary stent implantation. *JAMA*. 2010;303:754-762
32. Qayyum R, Becker DM, Yanek LR, Faraday N, Vaidya D, Mathias R, Kral BG, Becker LC. Greater collagen-induced platelet aggregation following cyclooxygenase 1 inhibition predicts incident acute coronary syndromes. *Clin Transl Sci*. 2015;8:17-22
33. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*. 1998;62:1198-1211



Impact of genetic variation in ANKRD26 (rs191015656) on ADP-induced platelet aggregation in discovery and replication cohorts. Optical aggregation was performed in platelet rich plasma in response to ADP. The concentrations of ADP used to induce aggregation (from left to right) were as follows: 0.725 uM, 3 uM, and 2 uM, in Caerphilly Prospective Study (CaPS), the Framingham Heart Study (FHS), and GeneSTAR (GS), respectively. Individual black points represent platelet values for heterozygotes for rs191015656.

220x220mm (184 x 184 DPI)

Table 1. Sample characteristics of the study cohorts

	Discovery Meta-Analysis Cohorts			Replication Cohorts	
	European Ancestry			European Ancestry	African American
	GeneSTAR N= 1,060	FHS N=2,604	PAPI N=549	Caerphilly N=1,184	GeneSTAR N=613
Age, mean (SD), yr	42.4 (12.6)	54.4 (9.7)	44.9 (13.8)	56.8 (4.4)	42.2 (12.2)
Female sex, %	551 (52.0)	1,404 (53.9)	282 (51.4)	0 (0)	381 (62.2)
BMI, mean (SD), kg/m2	28.7 (6.3)	27.4 (4.9)	27.3 (4.9)	26.7 (3.6)	32.2 (8.3)
HTN, %	263 (24.8)	509 (19.6)	19 (3.5)	668 (56.7)	229 (37.4)
Diabetes, %	56 (5.3)	161 (6.2)	4 (0.7)	34 (2.9)	67 (11.0)
LDL cholesterol, mean (SD), mg/dL	125.4 (37.4)	126.5 (32.8)	136.8 (44.3)	144.3 (35.5)	121.6 (38.0)
Current smoker, %	256 (24.2)	473 (18.2)	52 (9.5)	398 (41.2)	187 (30.5)

BMI= body mass index; HTN = hypertension; LDL= low density lipoprotein

Table 2. Significant association findings from gene-based meta-analysis in discovery cohorts.

gene	dbSNPid	chr:pos	protein effect***	N	MAF	Ref	Alt	Nref	Nhet	Nalt	Nhetfam	direction	beta	Pvalue
*ANKRD26**	gene-based analysis												0.6193	7.13E-7
	rs61745662	10:27324401	E992A or E993A	4047	0.00049	T	G	4043	4	0	4	++X	1.4682	0.0039
	rs146159734	10:27306678	C1419Y or C1420Y	4047	0.00210	C	T	4030	17	0	15	++X	0.6729	0.0057
	rs139949439	10:27328937	E777Q or E778Q	4047	0.00285	C	G	4016	23	0	16	--X	-0.5668	0.0071
	rs191015656	10:27381431	T181I	4047	0.00124	G	A	4037	10	0	8	++X	0.8476	0.0080
	rs141748831	10:27329099	S723R or S724R	4047	0.00210	T	G	4030	17	0	12	++X	0.5914	0.018
	rs41304587	10:27324372	E1002K or E1003K	4047	0.00198	C	T	4031	16	0	15	++X	0.4527	0.07
	rs79092403	10:27382639	N111A	4047	0.00025	T	C	4045	2	0	1	X+X	1.1681	0.12
	rs201638257	10:27375498	P227S	4047	0.00074	G	A	4039	6	0	3	++X	0.2835	0.51
	rs200100926	10:27382438	Q124L	4047	0.00012	T	A	4046	1	0	1	X-X	-0.5269	0.60
	rs201461870	10:27350084	V484A	4047	0.00012	A	G	4046	1	0	1	-XX	-0.5064	0.61
	rs199683454	10:27389253	M1I	4047	0.00012	C	T	4046	1	0	1	X+X	0.4922	0.62
	rs28477279	10:27335399	R622Q or R623Q	4047	0.00025	C	T	4045	2	0	2	X+X	0.1677	0.81
*ABCG1***	gene-based analysis												-2.8314	1.92E-6
	rs151254598	21:43708097	G358R or G369R	3794	0.00026	G	A	3792	2	0	2	--X	-2.7629	8.30E-5
	rs201983330	21:43708059	R345Q or R356Q	3794	0.00013	G	A	3793	1	0	1	-XX	-2.9617	0.0029
	rs138056067	21:43697009	S181L or S192L	3794	0.00066	C	T	3789	5	0	5	++X	0.2086	0.64

*Phenotype = ADP low 1; **Phenotype= ADP high 2; MAF= minor allele frequency; Ref= reference allele; Alt= alternate allele; Nhet= number of heterozygotes; Nhetfam= number of families with at least one heterozygote; direction of effect appears in order - GS/FHS/PAPI, X = variant not present in cohort; ***alternate protein isoforms for ANKRD26 and ABCG1 exist that affect some or all of the relative amino acid positions

Table 3: Translational effect and replication of gene-based association findings.

gene	dbSNPid	PolyPhen Score	Predicted effect on protein	SIFT score	Predicted effect	Study Phase/Cohort	N	REF	ALT	MAF	beta	Pvalue	Rsq
ANKRD26	rs61745662	0.723	Possibly damaging	0.02	deleterious	Discovery meta	4047	T	G	0.00049	1.4682	0.0039	
						CaPS	1028	T	G	0.00001	-85.7905	0.53	0.983
						GeneSTAR AA	574	T	G	0.06882	-0.0040	0.97	
	rs146159734	0.091	benign	0.73	tolerated	Discovery meta	4047	C	T	0.00210	0.6729	0.0057	
						CaPS	1028	C	T	0.00119	-0.5449	0.42	0.843
						GeneSTAR AA	574	C	T	0.00087	-0.8436	0.39	
	rs139949439	0.998	Possibly damaging	0.02	deleterious	Discovery meta	4047	C	G	0.00285	-0.5668	0.0071	
						CaPS	1028	C	G	0.00331	0.6620	0.11	0.826
						GeneSTAR AA	574	C	G	0.00087	0.4553	0.64	
	rs191015656	**0.904**	**Possibly damaging**	**0**	**deleterious**	**Discovery meta**	**4047**	**G**	**A**	**0.00124**	**0.8476**	**0.0080**	
						CaPS	**1028**	**G**	**A**	**0.00140**	**1.5643**	**0.011**	**0.921**
						GeneSTAR AA	574	NA	NA	NA	NA	NA	
	rs141748831	0.372	benign	0.03	deleterious	Discovery meta	4047	T	G	0.00210	0.5914	0.018	
						CaPS	1028	T	G	0.00284	-0.5188	0.23	0.949
						GeneSTAR AA	574	NA	NA	NA	NA	NA	
ABCG1	rs151254598	0	benign	0.59	tolerated	Discovery meta	3794	G	A	0.00026	-2.7629	8.3E-5	
						CaPS	1028	G	A	0.00002	-153.752	0.0016	0.007
						GeneSTAR AA	606	NA	NA	NA	NA	NA	
	rs201983330	0.001	benign	0.97	tolerated	Discovery meta	3794	G	A	0.00013	-2.9617	0.0029	
						CaPS	1028	NA	NA	NA	NA	NA	NA
						GeneSTAR AA	606	NA	NA	NA	NA	NA	

MAF= minor allele frequency; Rsq= imputation accuracy measure; NA = no information available (no allele carriers)
Significant replication findings are highlighted

Table 4: Significant novel locus findings and replication from single nucleotide variant analysis.

gene	chr:pos	dbSNPid	snp annotation	Trait (uM)	Phase	Study	N	AF	Nref	Nhet	Nalt	beta	pval	direction
HCP5	6:31421297	rs2263316	intronic *(overlaps CD34-CD41+CD42+ MK cell H3K9me3 + hypermethylation peaks)*	ADP low2*	Discovery	meta	3697	0.65	480	1614	1602	0.132	3.5E-7	+++
				ADP 2	Discovery	GS	943	0.63	128	443	371	0.090	0.12	+
				ADP EC50	Discovery	FHS	2205	0.63	322	1001	882	0.139	9.1E-6	+
				ADP 2	Discovery	PAPI	549	0.79	30	170	349	0.164	0.032	+
				ADP 0.725	**Replication[#]**	**CaPS**	**1028**	**0.66**	**127**	**462**	**439**	**0.084**	**0.044**	**+**
				ADP 2	**Replication[#]**	**AA**	**574**	**0.71**	**56**	**217**	**300**	**0.121**	**0.039**	**+**
TRIM24	7:138157390	rs1874326	intronic *(overlaps CD34-CD41+CD42+ MK cell H3K27ac, H3K4me1+ hypermethylation peaks)*	Epi high1*	Discovery	meta	2730	0.85	57	690	1983	-0.209	1.3E-7	---
				Epi 10	Discovery	GS	1050	0.86	12	261	777	-0.262	1.3E-4	-
				Epi 3	Discovery	FHS	1131	0.86	27	260	844	-0.170	4.9E-3	-
				Epi 10	Discovery	PAPI	549	0.82	18	169	362	-0.205	0.011	-
				Epi 10	Replication[#]	AA	607	0.85	9	168	430	0.001	0.49	+

highlighted denotes significant replication finding; *see Supplemental Table 1 for meta-analysis phenotype descriptions and mean (SD) in each cohort; [#]one-sided test statistics are reported since replication required *a priori* the same direction of effect as discovery; Epinephrine data was not acquired in CaPS and thus unavailable for replication of the TRIM24 signal

Supplemental Table 1: Harmonizing platelet phenotypes for meta-analysis

Agonist	Framingham Phenotype N=2,604	GeneSTAR Phenotype N=1,060	PAPI Phenotype N=549
ADP low 1 mean (SD)	Max aggregation to 3µM 67.4% (24.9%)	Max aggregation to 2µM 42.8% (26.2%)	Max aggregation to 2µM 39.1% (21.7%)
ADP high 1 mean (SD)	Max aggregation to 5µM 77.4% (18.7%)	Max aggregation to 10µM 78.4% (14.3%)	Max aggregation to 10µM 68.3% (12.6%)
ADP low 2 mean (SD)	Threshold dose for >50% aggregation 3.3 uM (1.5 uM)	Max aggregation to 2µM 42.8% (26.2%)	Max aggregation to 2µM 39.1% (21.7%)
ADP high 2 mean (SD)	Threshold dose for >50% aggregation 3.3 uM (1.5 uM)	Max aggregation to 10µM 78.4% (14.3%)	Max aggregation to 10µM 68.3% (12.6%)
Epi low 1 mean (SD)	Max aggregation to 1µM 57.1% (31.2%)	Max aggregation to 2µM 54.0% (33.7%)	Max aggregation to 10µM 62.3% (25.2%)
Epi high 1 mean (SD)	Max aggregation to 3µM 65.9% (28.5%)	Max aggregation to 10µM 69.7% (28.6%)	Max aggregation to 10µM 62.3% (25.2%)
Epi low 2 mean (SD)	Threshold dose for >50% aggregation 2.0 uM (2.9 uM)	Max aggregation to 2µM 54.0% (33.7%)	Max aggregation to 10µM 62.3% (25.2%)
Epi high 2 mean (SD)	Threshold dose for >50% aggregation 2.0 uM (2.9 uM)	Max aggregation to 10µM 69.7% (28.6%)	Max aggregation to 10µM 62.3% (25.2%)
Collagen group1 mean (SD)	Lag time to 190 µg/mL* 81.9 sec (19.9 sec)	Lag time to 1µg/mL** 156 sec (90.3 sec)	Lag time to 1µg/mL** 58.4 sec (23.9 sec)
Collagen group2 mean (SD)	Lag time to 190 µg/mL 81.9 sec (19.9 sec)	Lag time to 2µg/mL 106 sec (63.0 sec)	Lag time to 2µg/mL 49.2 sec (16.3 sec)
Collagen group3 mean (SD)	Lag time to 190 µg/mL 81.9 sec (19.9sec)	Lag time to 5µg/mL 71.9 sec (35.8 sec)	Lag time to 5µg/mL 38.5 sec (11.1 sec)
Collagen group4 mean (SD)	Lag time to 190 µg/mL 81.9 sec (19.9 sec)	Lag time to 10µg/mL 61.8 sec (23.8 sec)	Lag time to 10µg/mL 33.1 sec (10.1 sec)

*Collagen was obtained from Bio-Data Corp (Horsham, PA)

**Collagen was obtained from Chrono-Log Corp (Havertown, PA)

Supplemental Table 2. Gene-based analysis suggestive signals (P<1.0E-04).

phenotype	gene	N	pval	no. site	beta	se	direction.burden .by.study	direction.meta .single.var	test
ADP high 1	ABCG1	3275	3.49E-6	2	-2.417	0.498	--=	--	VT
Epi high 1	ZNF197	2730	7.06E-6	4	1.196	0.258	++=	++++	VT
Collagen 1	SPATA4	3753	1.02E-5	1	1.598	0.131	=+=	+	VT
Collagen 2	SPATA4	3753	1.02E-5	1	1.598	0.131	=+=	+	VT
Collagen 10	SPATA4	3742	1.02E-5	1	1.598	0.131	=+=	+	VT
Collagen 3	SPATA4	3753	1.02E-5	1	1.598	0.131	=+=	+	VT
Collagen 1	SPATA4	3753	1.02E-5	1	1.598	0.131	=+=	+	SKAT
Collagen 2	SPATA4	3753	1.02E-5	1	1.598	0.131	=+=	+	SKAT
Collagen 10	SPATA4	3742	1.02E-5	1	1.598	0.131	=+=	+	SKAT
Collagen 3	SPATA4	3753	1.02E-5	1	1.598	0.131	=+=	+	SKAT
ADP high 1	TPR	3275	1.22E-5	10	-0.019	0.119	+-=	-+--+----+	SKAT
ADP high 2	ZNF610	3794	1.58E-5	5	-0.292	0.086	-+-	--++-	SKAT
ADP low 2	ZNF25	3697	1.87E-5	1	2.963	0.479	+==	+	VT
ADP low 2	ZNF25	3697	1.87E-5	1	2.963	0.479	+==	+	SKAT
Epi high 1	LAPTM4B	2730	1.87E-5	5	1.356	0.306	++=	+++++	VT
ADP low 1	ANKRD26	4047	2.07E-5	12	0.343	0.103	-+=	+++-++-++-++	SKAT
Collagen 10	COL4A2	3742	2.28E-5	11	-0.708	0.159	--=	----------	VT
ADP low 1	ABCC9	4047	2.39E-5	4	1.629	0.363	++=	++++	VT
Collagen 1	HAUS3	3753	3.68E-5	3	-1.623	0.364	--=	---	VT
ADP low 1	MICU1	4047	4.06E-5	2	-1.538	0.382	--=	--	SKAT
Epi high 1	SLCO2B1	2730	4.09E-5	7	-0.369	0.285	-+=	++----+	SKAT
Epi high 1	ZNF197	2730	4.63E-5	4	1.196	0.258	++=	++++	SKAT
ADP high 1	ZNF610	3275	5.61E-5	6	-0.320	0.089	-+-	--+++-	SKAT
ADP high 2	LEFTY1	3794	6.13E-5	1	1.422	0.126	++=	+	VT
ADP high 2	LEFTY1	3794	6.13E-5	1	1.422	0.126	++=	+	SKAT
Epi high 1	RLBP1	2730	6.85E-5	3	0.969	0.272	+++	-++	SKAT
Collagen 1	SLC6A16	3753	7.04E-5	13	0.329	0.108	+++	+++--+--+--++	SKAT
Epi low 2	ASIC5	3797	7.21E-5	9	-0.898	0.211	+-=	-------+-	VT
ADP high 1	DDX60	3275	7.52E-5	8	0.584	0.175	++=	++++--++	SKAT
Epi high 2	RLBP1	3798	8.20E-5	3	0.981	0.238	+++	+++	VT
ADP high 1	RAB25	3275	8.25E-5	3	-1.118	0.299	=-=	+--	SKAT
Epi high 1	TXNIP	2730	8.39E-5	1	-2.891	0.706	=-=	-	VT
Epi low 2	DYNC2H1	3797	8.52E-5	20	0.821	0.187	++=	-+-+++++++-+-+++-++	VT
ADP high 2	GJA9	3794	8.90E-5	3	1.734	0.414	++=	+++	VT

Epi high 1	DIRC1	2730	8.91E-5	3	-0.558	0.137	---	---	SKAT
Epi high 1	RLBP1	2730	9.32E-5	2	1.198	0.295	+++	++	VT
ADP low 1	SLC6A17	4047	9.54E-5	3	2.343	0.577	=+=	+++	VT
Epi low 2	LIM2	3797	9.60E-5	3	1.199	0.289	++=	+++	VT

Supplemental Table 3. Linkage disequilibrium calculations among *ANKRD26* variants in unrelated individuals in FHS.

SNP1	SNP2	D'	r2	FisherExact.P
rs146159734	rs41304587	1	5.64E-06	1
rs146159734	rs61745662	1	4.93E-06	1
rs146159734	rs139949439	1	3.52E-06	1
rs146159734	rs141748831	1	2.11E-06	1
rs146159734	rs28477279	1	7.03E-07	1
rs146159734	rs201461870	1	7.03E-07	1
rs146159734	rs201638257	1	4.93E-06	1
rs146159734	rs191015656	1	7.03E-07	1
rs146159734	rs200100926	NA	NA	1
rs146159734	rs79092403	1	7.03E-07	1
rs41304587	rs61745662	1	6.58E-06	1
rs41304587	rs139949439	1	4.70E-06	1
rs41304587	rs141748831	1	2.82E-06	1
rs41304587	rs28477279	1	9.39E-07	1
rs41304587	rs201461870	1	9.39E-07	1
rs41304587	rs201638257	1	6.58E-06	1
rs41304587	rs191015656	1	9.39E-07	1
rs41304587	rs200100926	NA	NA	1
rs41304587	rs79092403	1	9.39E-07	1
rs61745662	rs139949439	1	4.11E-06	1
rs61745662	rs141748831	1	2.46E-06	1
rs61745662	rs28477279	1	8.21E-07	1
rs61745662	rs201461870	1	8.21E-07	1
rs61745662	rs201638257	1	5.76E-06	1
rs61745662	rs191015656	1	8.21E-07	1
rs61745662	rs200100926	NA	NA	1
rs61745662	rs79092403	1	8.21E-07	1
rs139949439	rs141748831	1	1.76E-06	1
rs139949439	rs28477279	1	5.86E-07	1
rs139949439	rs201461870	1	5.86E-07	1
rs139949439	rs201638257	1	4.11E-06	1
rs139949439	rs191015656	1	5.86E-07	1
rs139949439	rs200100926	NA	NA	1
rs139949439	rs79092403	1	5.86E-07	1
rs141748831	rs28477279	1	3.51E-07	1
rs141748831	rs201461870	1	3.51E-07	1
rs141748831	rs201638257	1	2.46E-06	1
rs141748831	rs191015656	1	3.51E-07	1
rs141748831	rs200100926	NA	NA	1
rs141748831	rs79092403	1	3.51E-07	1

rs28477279	rs201461870	1	1.17E-07	1
rs28477279	rs201638257	1	8.21E-07	1
rs28477279	rs191015656	1	1.17E-07	1
rs28477279	rs200100926	NA	NA	1
rs28477279	rs79092403	1	1.17E-07	1
rs201461870	rs201638257	1	8.21E-07	1
rs201461870	rs191015656	1	1.17E-07	1
rs201461870	rs200100926	NA	NA	1
rs201461870	rs79092403	1	1.17E-07	1
rs201638257	rs191015656	1	8.21E-07	1
rs201638257	rs200100926	NA	NA	1
rs201638257	rs79092403	1	8.21E-07	1
rs191015656	rs200100926	NA	NA	1
rs191015656	rs79092403	1	1.17E-07	1
rs200100926	rs79092403	NA	NA	1

Supplemental Table 4. SNV known or suggestive signals (P<7.0E-06).

phenotype	dbSNPID	chr:pos	SKATgene	snp annotation	Ref	Alt	N	EAF	beta	se	direction	pval
Epi low 2	rs12566888	1:156869047	*PEAR1*	intronic	G	T	3797	0.094	-0.291	0.040	---	2.98E-13
ADP low 1	rs12566888	1:156869047	*PEAR1*	intronic	G	T	4047	0.098	-0.271	0.038	---	1.26E-12
ADP low 2	rs12566888	1:156869047	*PEAR1*	intronic	G	T	3697	0.095	-0.283	0.040	---	2.53E-12
Epi hgh 2	rs12566888	1:156869047	*PEAR1*	intronic	G	T	3798	0.094	-0.275	0.040	---	6.15E-12
Epi low 1	rs12566888	1:156869047	*PEAR1*	intronic	G	T	3607	0.096	-0.269	0.041	---	3.27E-11
Collagen 1	rs1671152	19:55526345	*GP6*	nonsynonymous	T	G	3753	0.835	-0.199	0.033	--+	1.01E-09
Collagen 2	rs1671152	19:55526345	*GP6*	nonsynonymous	T	G	3753	0.835	-0.198	0.033	---	1.44E-09
Collagen 4	rs1671152	19:55526345	*GP6*	nonsynonymous	T	G	3742	0.835	-0.194	0.033	--+	2.71E-09
ADP high 2	rs12566888	1:156869047	*PEAR1*	intronic	G	T	3794	0.095	-0.232	0.040	---	5.84E-09
Epi low 1	rs4545476	10:112922409	*exm2249039 (nearest ADRA2A)*	intergenic	C	T	3607	0.375	-0.144	0.025	---	1.44E-08
Collagen 1	rs1613662	19:55536595	*GP6*	nonsynonymous	G	A	3753	0.832	-0.181	0.032	--+	1.82E-08
Collagen 3	rs1671152	19:55526345	*GP6*	nonsynonymous	T	G	3753	0.835	-0.181	0.032	--+	2.33E-08
Collagen 2	rs1654413	19:55526359	*GP6*	nonsynonymous	A	T	3753	0.820	-0.173	0.032	---	4.55E-08
Collagen 1	rs1654413	19:55526359	*GP6*	nonsynonymous	A	T	3753	0.820	-0.170	0.031	--+	6.25E-08
Collagen 2	rs2304167	19:55527081	*GP6*	nonsynonymous	C	T	3753	0.820	-0.170	0.032	---	6.66E-08
Epi low 1	rs869244	10:112909105	*exm-rs869244 (nearest ADRA2A)*	intergenic	G	A	3607	0.330	-0.140	0.026	---	7.59E-08
Collagen 2	rs1613662	19:55536595	*GP6*	nonsynonymous	G	A	3753	0.832	-0.174	0.032	---	7.84E-08
Epi hgh 2	rs10761731	10:65027610	*JMJD1C*	intronic	A	T	3798	0.428	0.129	0.024	++-	9.72E-08
Epi low 2	rs869244	10:112909105	*exm-rs869244 (nearest ADRA2A)*	intergenic	G	A	3797	0.331	-0.135	0.025	---	9.77E-08
Collagen 1	rs2304167	19:55527081	*GP6*	nonsynonymous	C	T	3753	0.820	-0.167	0.031	--+	1.06E-07
Collagen 4	rs1654413	19:55526359	*GP6*	nonsynonymous	A	T	3742	0.820	-0.165	0.031	--+	1.51E-07

Collagen 4	rs1613662	19:55536595	*GP6*	nonsynonymous	G	A	3742	0.832	-0.169	0.032	--+	1.58E-07
Collagen 4	rs2304167	19:55527081	*GP6*	nonsynonymous	C	T	3742	0.820	-0.162	0.031	--+	2.72E-07
Epi hgh 2	rs342293	7:106372219	*exm-rs342293 (nearest PIK3CG)*	intergenic	C	G	3798	0.435	-0.121	0.024	---	3.24E-07
Epi low 2	rs342293	7:106372219	*exm-rs342293 (nearest PIK3CG)*	intergenic	C	G	3797	0.435	-0.121	0.024	---	3.33E-07
Epi low 2	rs10761731	10:65027610	*JMJD1C*	intronic	A	T	3797	0.428	0.123	0.024	++-	3.35E-07
Epi low 2	rs4545476	10:112922409	*exm2249039 (nearest ADRA2A)*	intergenic	C	T	3797	0.377	-0.126	0.025	---	3.78E-07
Collagen 3	rs1613662	19:55536595	*GP6*	nonsynonymous	G	A	3753	0.832	-0.161	0.032	--+	5.29E-07
Epi high 1	rs4909945	11:10673739	*MRVI1*	nonsynonymous	T	C	2730	0.659	-0.149	0.030	---	6.28E-07
Collagen 3	rs1654413	19:55526359	*GP6*	nonsynonymous	A	T	3753	0.820	-0.155	0.031	--+	7.45E-07
Epi low 1	rs2914908	5:117982046	*exm2270115 (nearest RPL11-2N5.2)*	intergenic	C	T	3607	0.490	0.120	0.024	+++	7.57E-07
Epi hgh 2	rs4909945	11:10673739	*MRVI1*	nonsynonymous	T	C	3798	0.665	-0.126	0.026	---	7.87E-07
Epi high 1	rs7940646	11:10669228	*MRVI1*	intronic	T	C	2730	0.658	-0.148	0.030	---	7.88E-07
ADP high 2	rs6943029	7:155762157	*exm-rs6943029 (nearest SHH)*	intergenic	G	A	3794	0.072	0.222	0.046	+++	1.13E-06
ADP high 2	rs2263316	6:31421297	*HCP5*	intergenic	A	G	3794	0.650	0.123	0.025	+++	1.15E-06
Collagen 3	rs2304167	19:55527081	*GP6*	nonsynonymous	C	T	3753	0.820	-0.152	0.031	--+	1.20E-06
Epi hgh 2	rs7940646	11:10669228	*MRVI1*	intronic	T	C	3798	0.664	-0.124	0.025	---	1.25E-06
Epi hgh 2	rs10761779	10:65274927	*JMJD1C*	intergenic	A	G	3798	0.500	0.115	0.024	+++	1.82E-06
ADP high 1	rs7940646	11:10669228	*MRVI1*	intronic	T	C	3275	0.662	-0.131	0.027	---	1.84E-06
Epi hgh 2	rs12355784	10:65121565	*JMJD1C*	intronic	C	A	3798	0.500	0.113	0.024	+++	2.54E-06
Epi hgh 2	rs869244	10:112909105	*exm-rs869244 (nearest ADRA2A)*	intergenic	G	A	3798	0.330	-0.120	0.025	---	2.61E-06

Epi hgh 2	rs4545476	10:112922409	*exm2249039 (nearest ADRA2A)*	intergenic	C	T	3798	0.377	-0.116	0.025	---	2.61E-06
Epi low 2	rs10761779	10:65274927	*JMJD1C*	intergenic	A	G	3797	0.501	0.113	0.024	+++	2.67E-06
ADP high 1	rs4909945	11:10673739	*MRVI1*	nonsynonymous	T	C	3275	0.662	-0.129	0.027	---	2.72E-06
ADP high 1	rs1052763	22:19119751	*TSSK2*	nonsynonymous	C	T	3275	0.183	-0.154	0.033	---	3.08E-06
Epi low 2	rs4909945	11:10673739	*MRVI1*	nonsynonymous	T	C	3797	0.665	-0.118	0.025	---	3.71E-06
Epi low 2	rs12355784	10:65121565	*JMJD1C*	intronic	C	A	3797	0.500	0.110	0.024	+++	4.26E-06
Collagen 3	rs7832232	8:38469303	*exm-rs7832232 (nearest RNF5P1)*	intergenic	A	G	3753	0.474	0.109	0.024	+++	4.48E-06
Collagen 4	rs7832232	8:38469303	*exm-rs7832232 (nearest RNF5P1)*	intergenic	A	G	3742	0.474	0.109	0.024	+++	4.70E-06
Epi low 1	rs144150699	3:54952554	*LRTM1*	nonsynonymous	G	C	3058	0.001	2.073	0.455	=++	5.26E-06
ADP low 1	rs4909945	11:10673739	*MRVI1*	nonsynonymous	T	C	4047	0.668	-0.113	0.025	---	5.81E-06
ADP low 1	rs7940646	11:10669228	*MRVI1*	intronic	T	C	4047	0.667	-0.112	0.025	---	6.28E-06
Epi low 2	rs7940646	11:10669228	*MRVI1*	intronic	T	C	3797	0.664	-0.115	0.025	---	6.29E-06
Epi low 2	rs115780313	1:34254272	*CSMD2*	nonsynonymous	T	C	3797	0.002	1.300	0.288	++=	6.45E-06
Collagen 1	rs9267673	6:31883679	*C2*	intronic	C	T	3753	0.114	-0.173	0.039	---	6.88E-06

Supplemental Table 5. SNV associations confirming prior reported genome-wide significant findings[4].

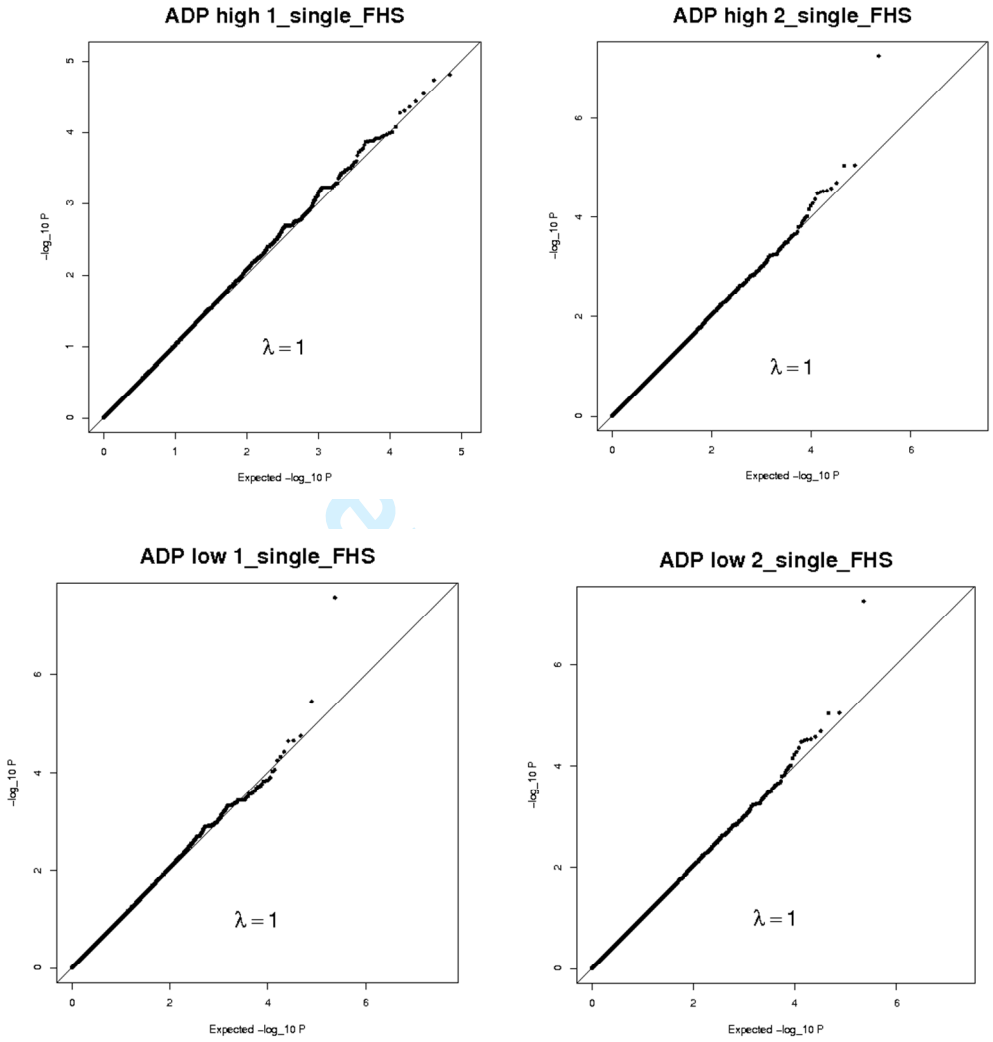
gene/region	chr:pos	dbSNPid	translational effect	Phenotype	N	AF	beta	p-value	direction
PEAR1	1:156869047	rs12566888	intronic	ADP low1	4047	0.098	-0.2708	1.26E-12	---
				ADP low2	3697	0.095	-0.2828	2.53E-12	---
				ADP high2	3794	0.095	-0.2317	5.84E-09	---
				Epi low1	3607	0.096	-0.2688	3.27E-11	---
				Epi low2	3797	0.094	-0.2913	2.98E-13	---
				Epi high2	3798	0.094	-0.2751	6.15E-12	---
JMJD1C	10:65027610	rs10761731	intronic	Epi low2	3797	0.428	0.1231	3.35E-07	++-
				Epi high2	3798	0.428	0.1290	9.72E-08	++-
ADRA2A	10:112922409	rs4545476	intergenic	Epi low1	3607	0.375	-0.1437	1.44E-08	---
				Epi low2	3797	0.377	-0.1256	3.78E-07	---
ADRA2A	10:112909105	rs869244	intergenic	Epi low1	3607	0.330	-0.1400	7.59E-08	---
				Epi low2	3797	0.331	-0.1354	9.77E-08	---
GP6	19:55526345	rs1671152	nonsynonymous	Collagen 1	3753	0.835	-0.1988	1.01E-09	--+
				Collagen 2	3753	0.835	-0.1979	1.44E-09	---
				Collagen 5	3753	0.835	-0.1810	2.33E-08	--+
				Collagen 10	3742	0.835	-0.1938	2.71E-09	--+
PIK3CG region	7:106372219	rs342293	intergenic	Epi high2	3798	0.435	-0.1212	3.24E-07	---
				Epi low2	3797	0.435	-0.1208	3.33E-07	---

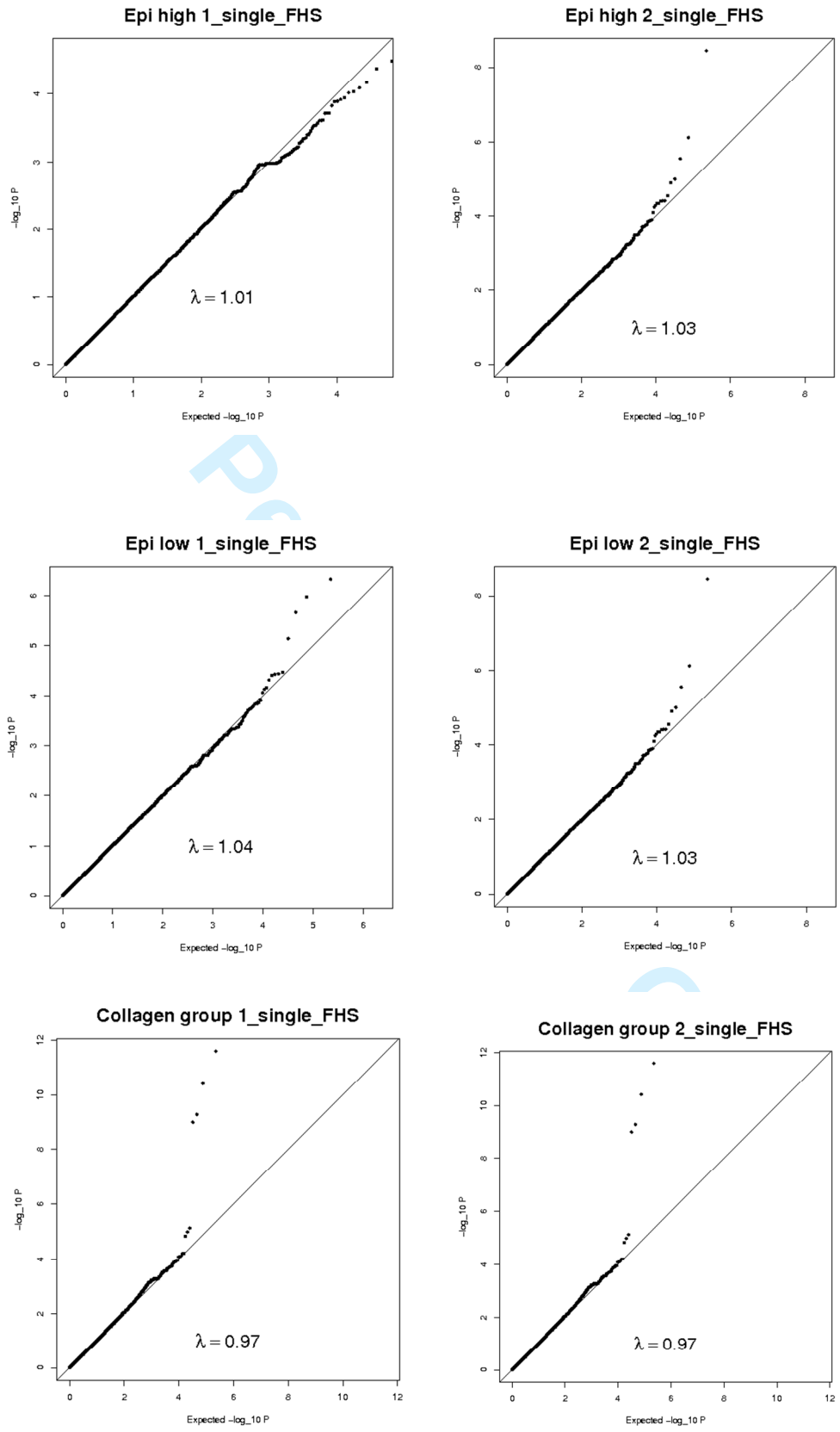
Supplemental Table 6. Power analysis for 4,000 individuals with family structure. We used SOLAR[32] to simulate phenotypes conditional on the observed family structures in the FHS sample. The phenotypes were simulated to have variance of 1 and polygenic heritability of 0.25. We considered the QTL (additive model) with MAF of 0.01, 0.05, 0.1, 0.2, 0.3 and 0.4, and QTL variance of 0.0025, 0.005, 0.0075 and 0.01. 1,000 of replicates were used to estimate power for each considered scenario. Simulated phenotypes were analyzed using a linear mixed effects model implemented in the *lmekin* function from the *coxme* R package where a relationship coefficient matrix was used to account for familial correlation.

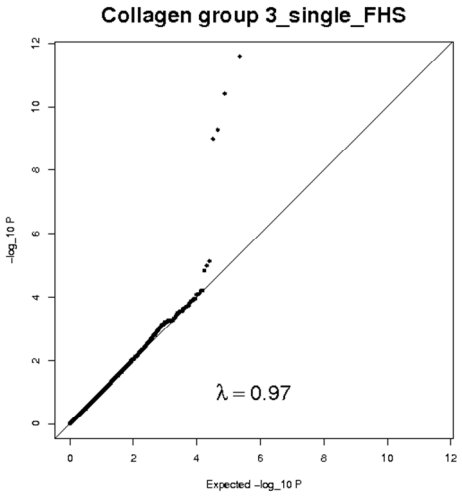
maf	qtl_var	power
0.01	0.0025	0.03
0.01	0.005	0.25
0.01	0.0075	0.577
0.01	0.01	0.813
0.05	0.0025	0.03
0.05	0.005	0.243
0.05	0.0075	0.595
0.05	0.01	0.848
0.1	0.0025	0.023
0.1	0.005	0.232
0.1	0.0075	0.601
0.1	0.01	0.863
0.2	0.0025	0.022
0.2	0.005	0.23
0.2	0.0075	0.598
0.2	0.01	0.843
0.3	0.0025	0.02
0.3	0.005	0.227
0.3	0.0075	0.594
0.3	0.01	0.861
0.4	0.0025	0.027
0.4	0.005	0.244
0.4	0.0075	0.586
0.4	0.01	0.856

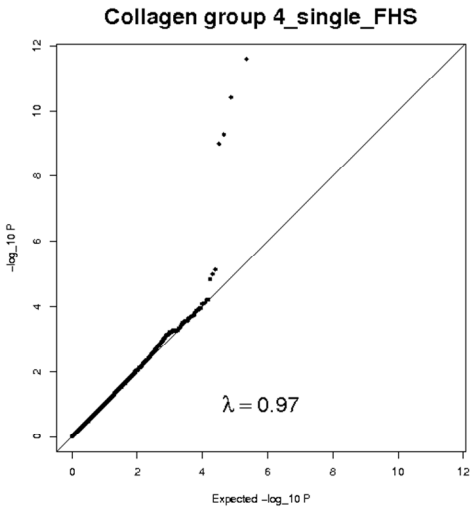
Supplemental Figure 1. Cohort level and meta-analysis QQ plots and genomic control values for single SNP and gene-based tests.

1. Framingham Heart Study

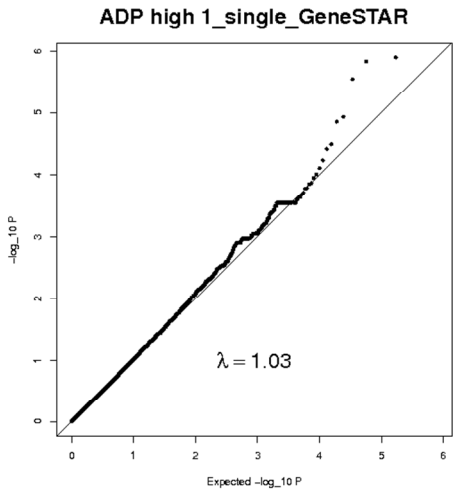


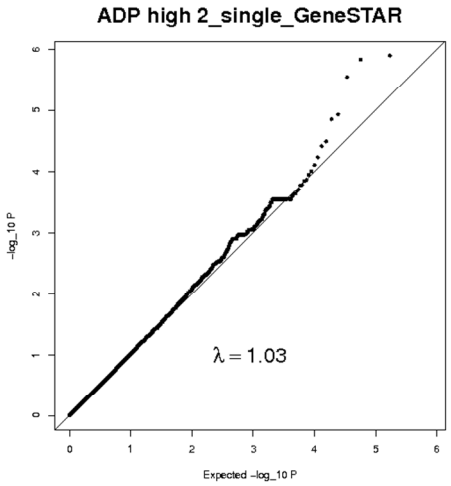

Epi high 1_single_FHS
λ = 1.01
Epi high 2_single_FHS
λ = 1.03
Epi low 1_single_FHS
λ = 1.04
Epi low 2_single_FHS
λ = 1.03
Collagen group 1_single_FHS
λ = 0.97
Collagen group 2_single_FHS
λ = 0.97
Expected −log_10 P
−log_10 P

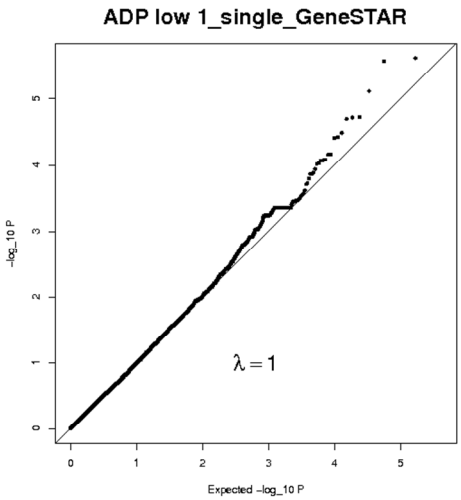

Collagen group 3_single_FHS
λ = 0.97
−log_10 P
Expected −log_10 P

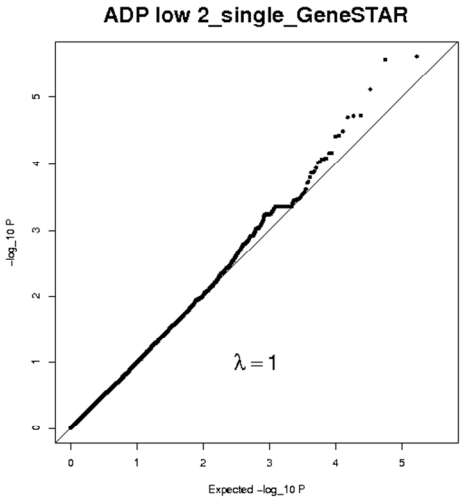

Collagen group 4_single_FHS
λ = 0.97
−log_10 P
Expected −log_10 P

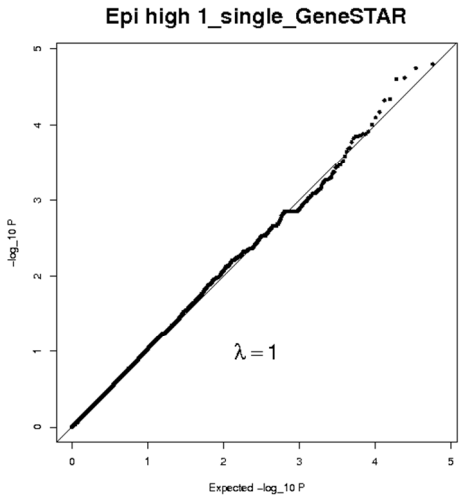
2. GeneSTAR

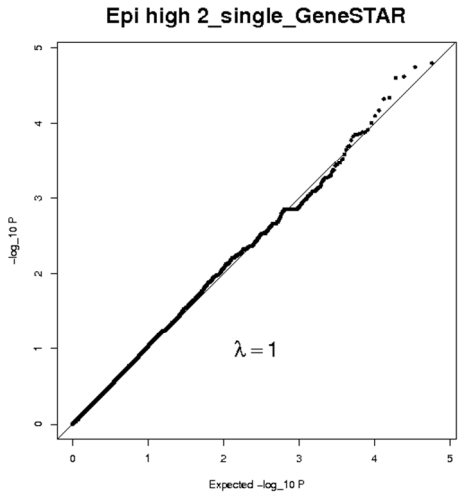

ADP high 1_single_GeneSTAR
λ = 1.03
−log_10 P
Expected −log_10 P

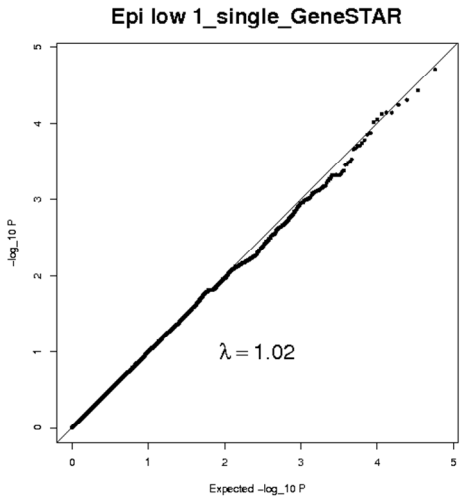

ADP high 2_single_GeneSTAR
λ = 1.03
−log_10 P
Expected −log_10 P

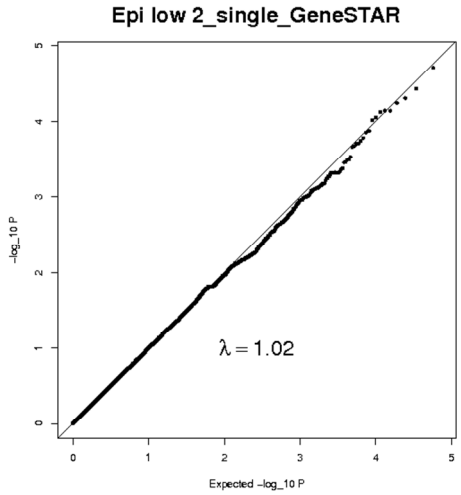

ADP low 1_single_GeneSTAR
λ = 1
−log_10 P
Expected −log_10 P

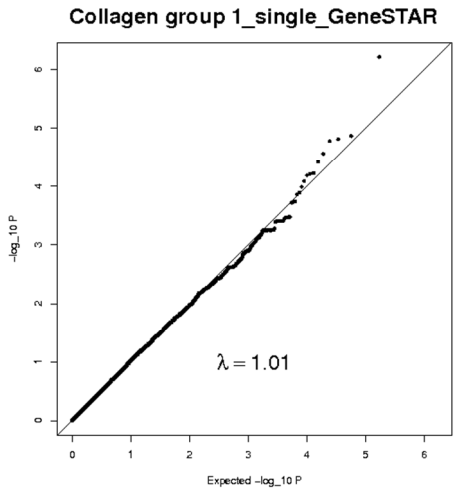

ADP low 2_single_GeneSTAR
λ = 1
−log_10 P
Expected −log_10 P

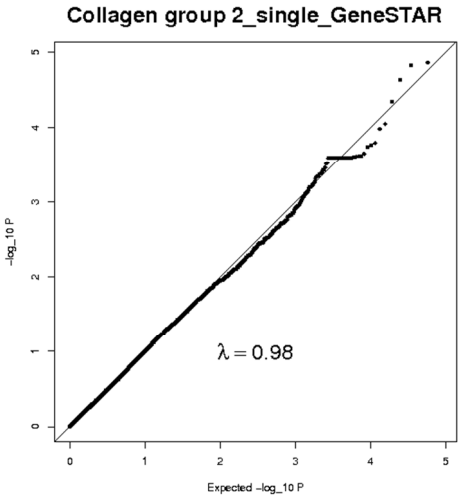

Epi high 1_single_GeneSTAR
λ = 1
Expected −log_10 P
−log_10 P

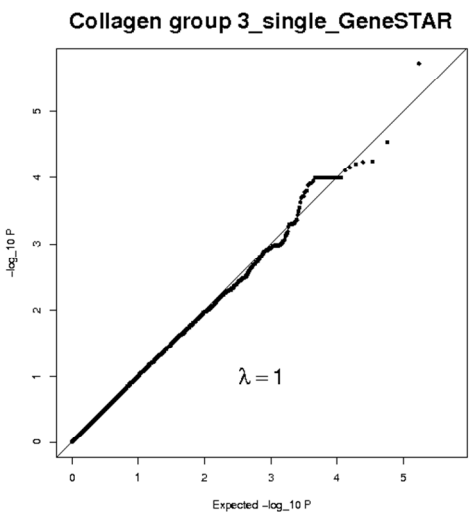

Epi high 2_single_GeneSTAR
λ = 1
Expected −log_10 P
−log_10 P

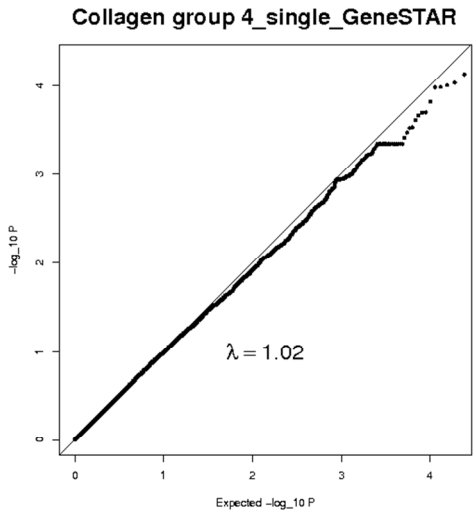

Epi low 1_single_GeneSTAR
λ = 1.02
Expected −log_10 P
−log_10 P


Epi low 2_single_GeneSTAR
λ = 1.02
Expected −log_10 P
−log_10 P

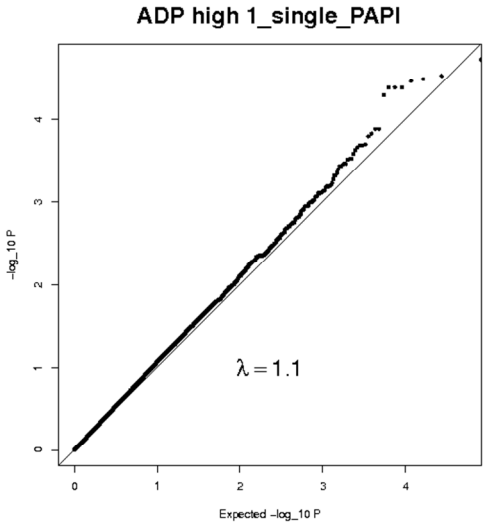

Collagen group 1_single_GeneSTAR
λ = 1.01
Expected −log_10 P
−log_10 P

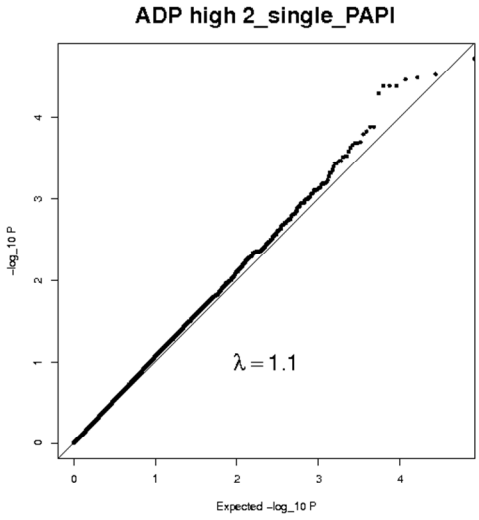

Collagen group 2_single_GeneSTAR
λ = 0.98
Expected −log_10 P
−log_10 P

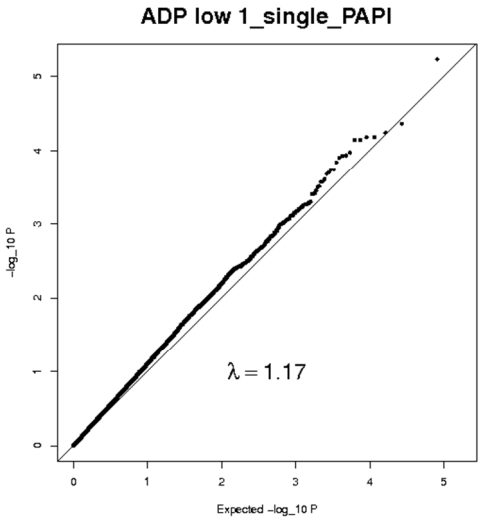

Collagen group 3_single_GeneSTAR
λ = 1
Expected −log_10 P
−log_10 P

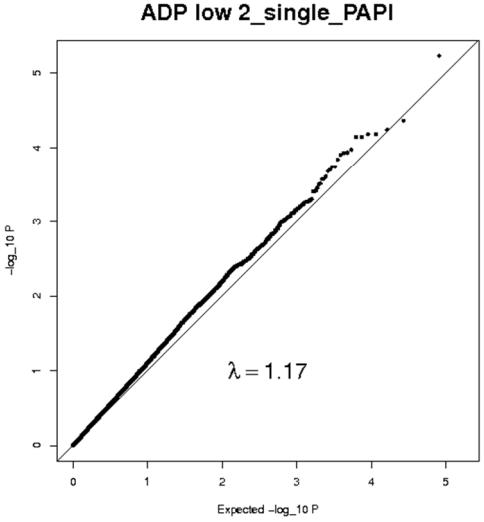

Collagen group 4_single_GeneSTAR
λ = 1.02
Expected −log_10 P
−log_10 P

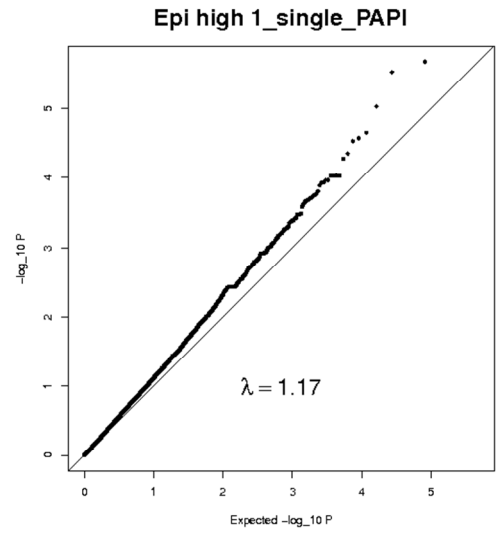
3. PAPI

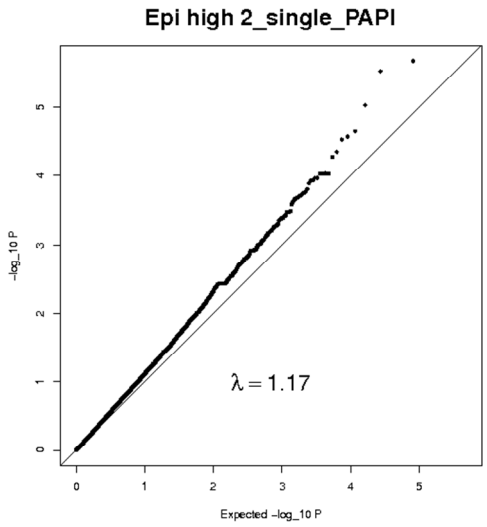

ADP high 1_single_PAPI
λ = 1.1
Expected −log_10 P
−log_10 P

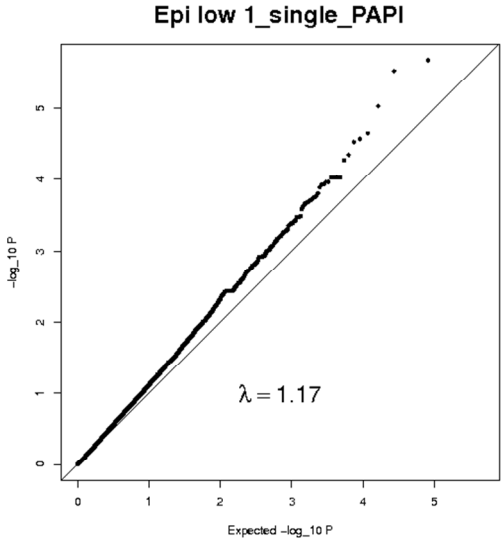

ADP high 2_single_PAPI
λ = 1.1
Expected −log_10 P
−log_10 P

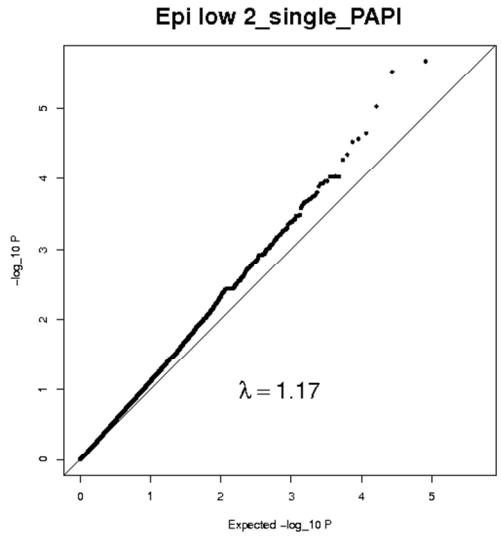

ADP low 1_single_PAPI
λ = 1.17
Expected −log_10 P
−log_10 P

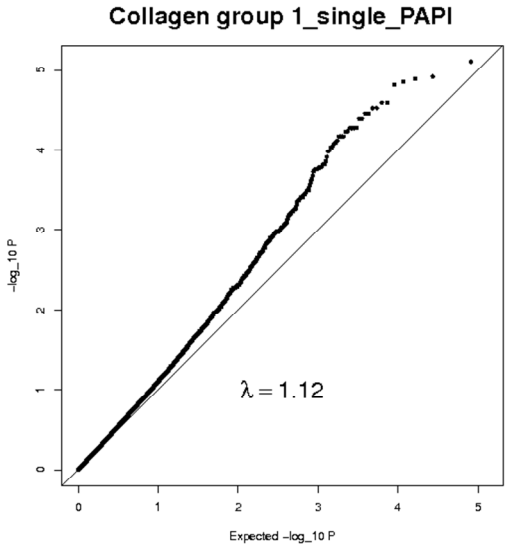

ADP low 2_single_PAPI
λ = 1.17
Expected −log_10 P
−log_10 P

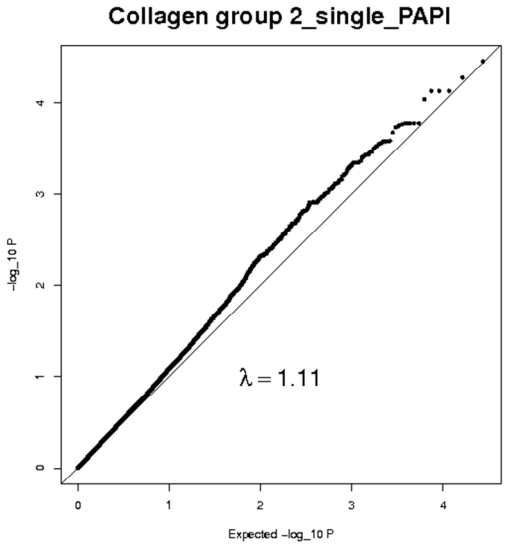

Epi high 1_single_PAPI
λ = 1.17
−log_10 P
Expected −log_10 P

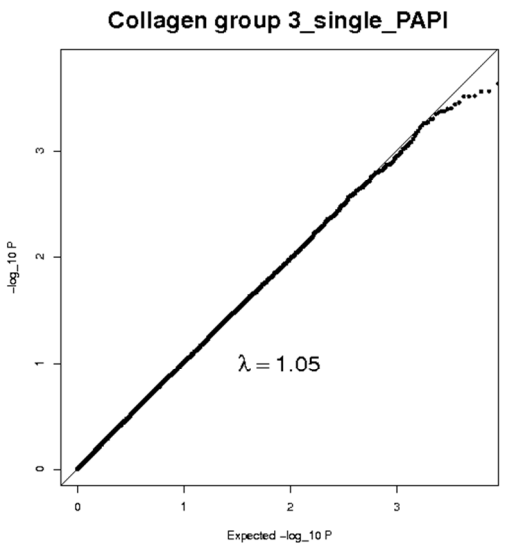

Epi high 2_single_PAPI
λ = 1.17
−log_10 P
Expected −log_10 P

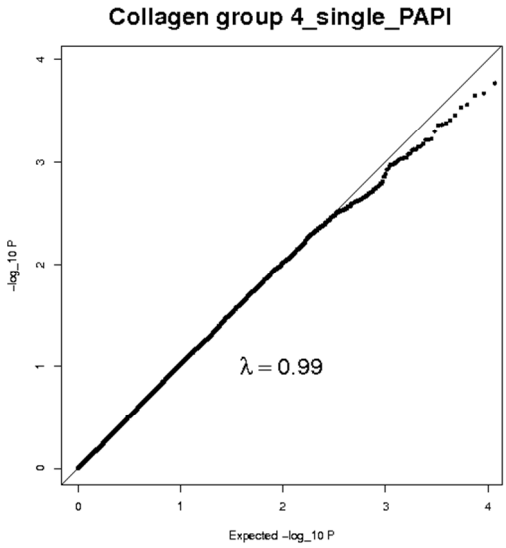

Epi low 1_single_PAPI
λ = 1.17
−log_10 P
Expected −log_10 P


Epi low 2_single_PAPI
λ = 1.17
−log_10 P
Expected −log_10 P

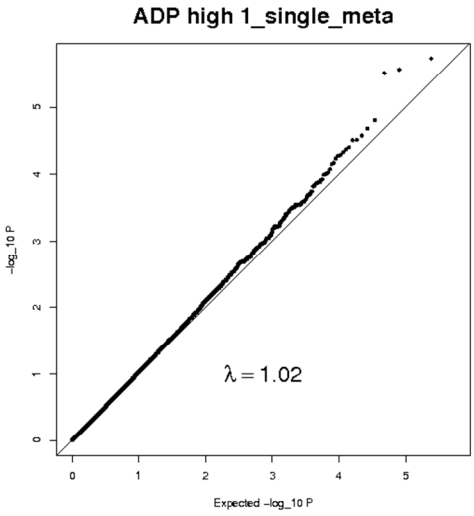

Collagen group 1_single_PAPI
λ = 1.12
−log_10 P
Expected −log_10 P

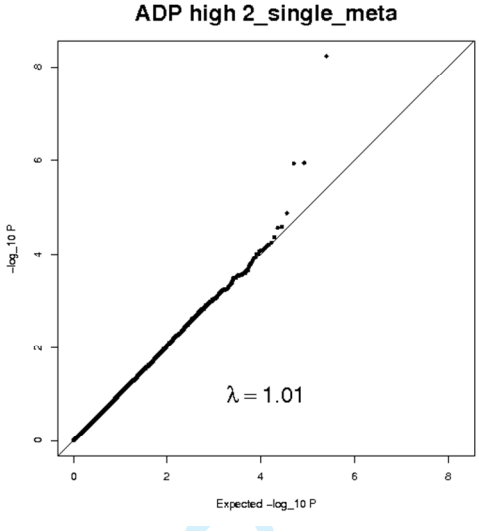

Collagen group 2_single_PAPI
λ = 1.11
−log_10 P
Expected −log_10 P

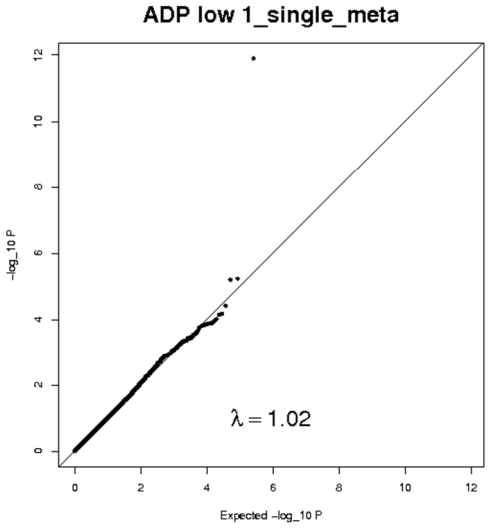


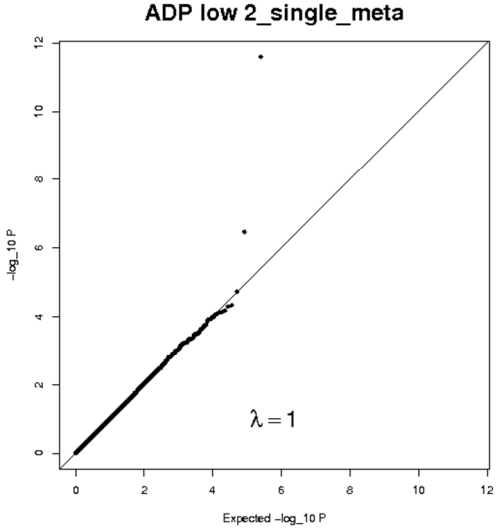


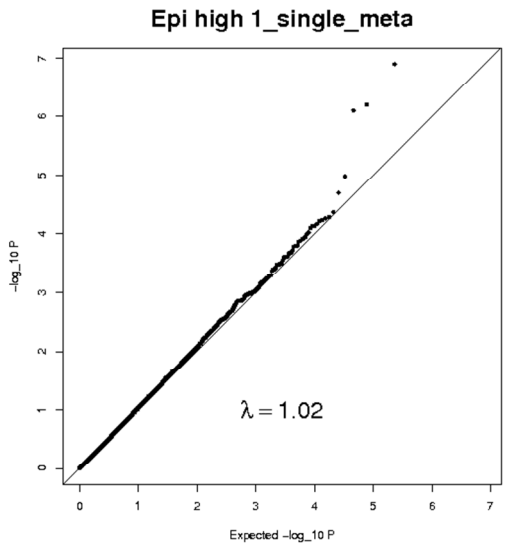
4. Single SNP meta-analysis

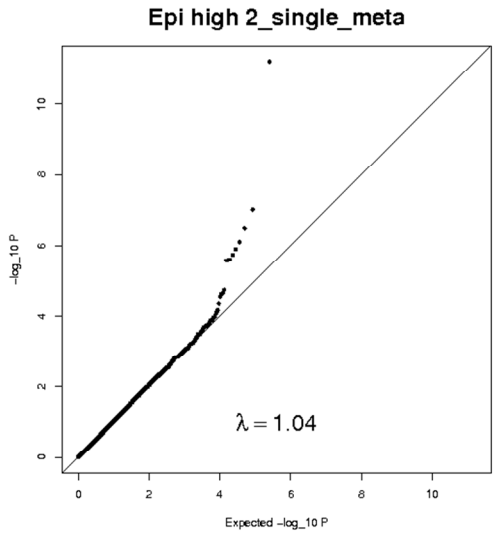


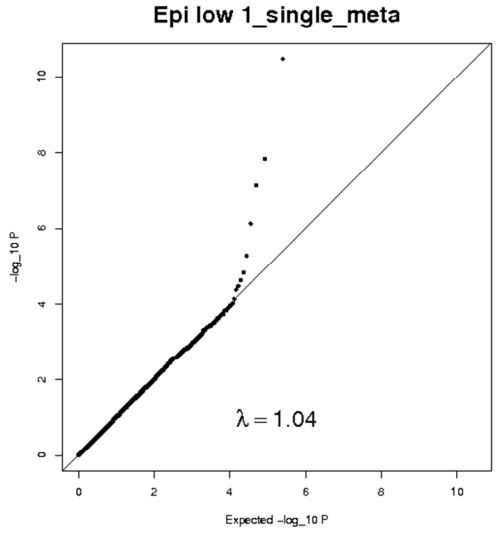


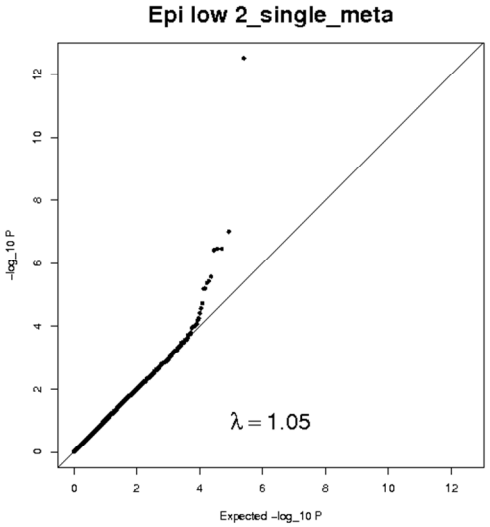


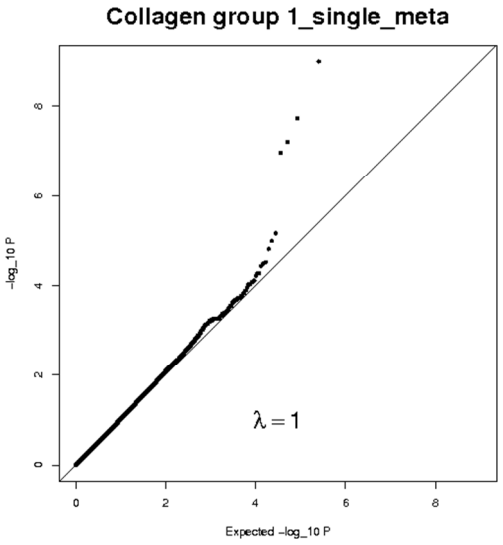


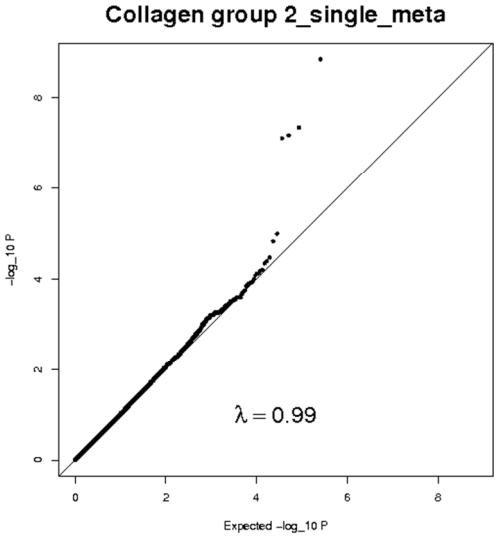

Epi high 1_single_meta
λ = 1.02
−log_10 P
Expected −log_10 P

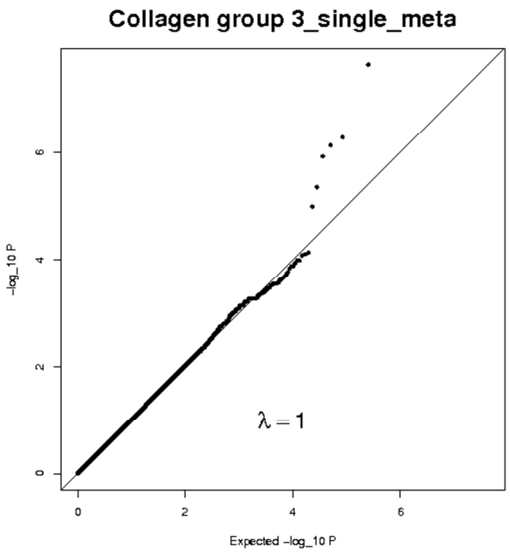

Epi high 2_single_meta
λ = 1.04
−log_10 P
Expected −log_10 P

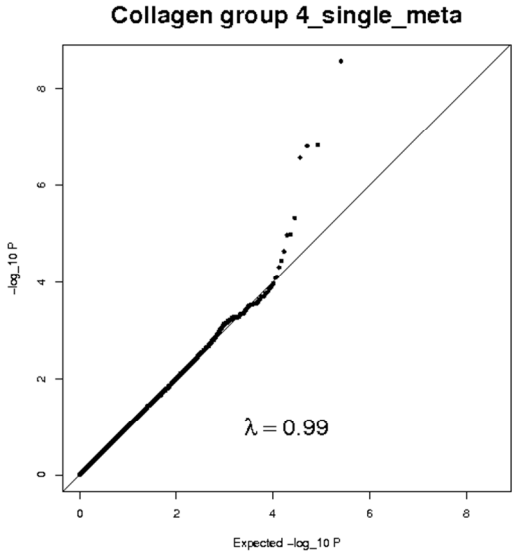

Epi low 1_single_meta
λ = 1.04
−log_10 P
Expected −log_10 P


Epi low 2_single_meta
λ = 1.05
−log_10 P
Expected −log_10 P

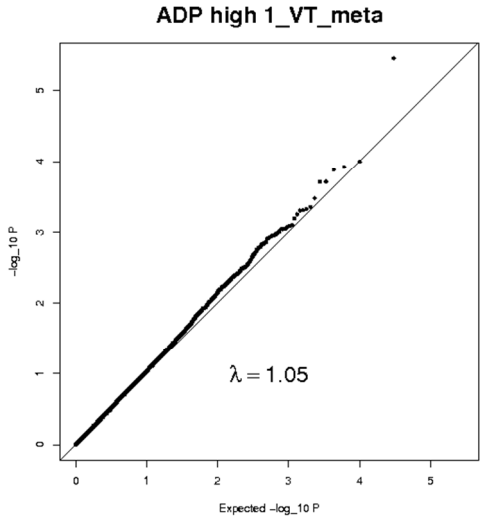

Collagen group 1_single_meta
λ = 1
−log_10 P
Expected −log_10 P

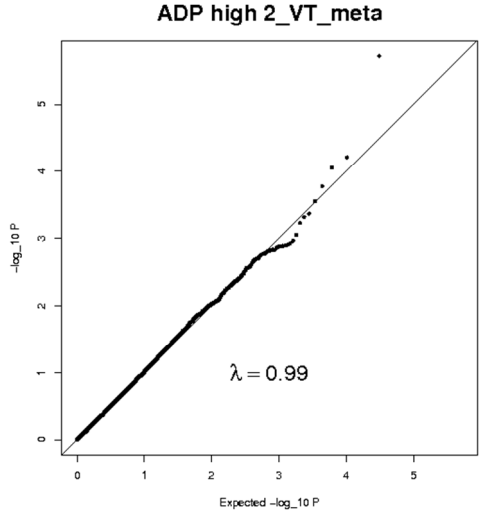

Collagen group 2_single_meta
λ = 0.99
−log_10 P
Expected −log_10 P

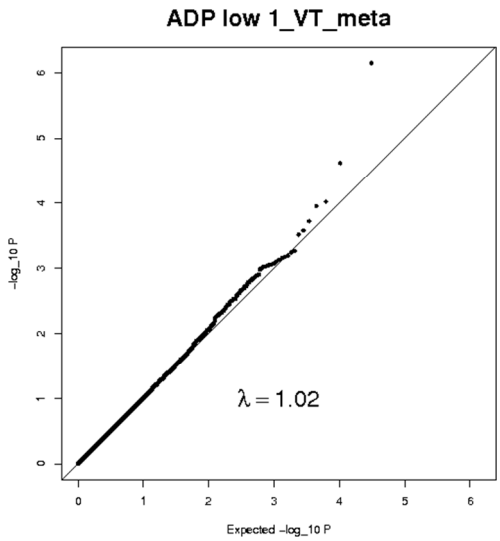


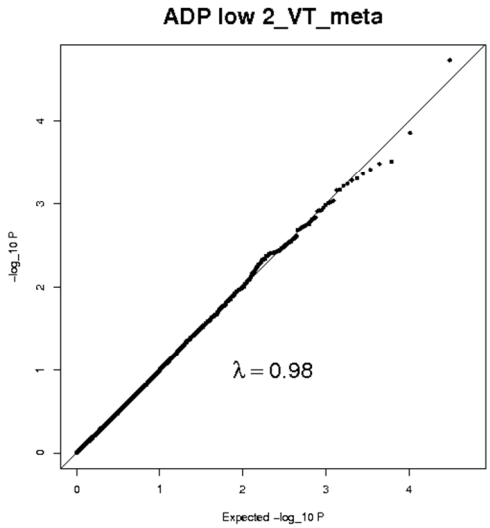


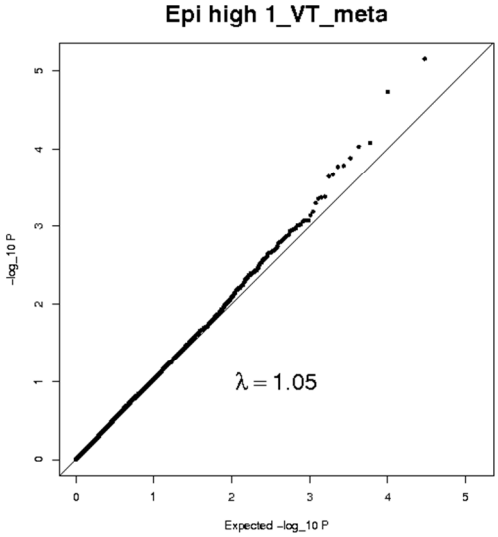
5. VT meta-analysis

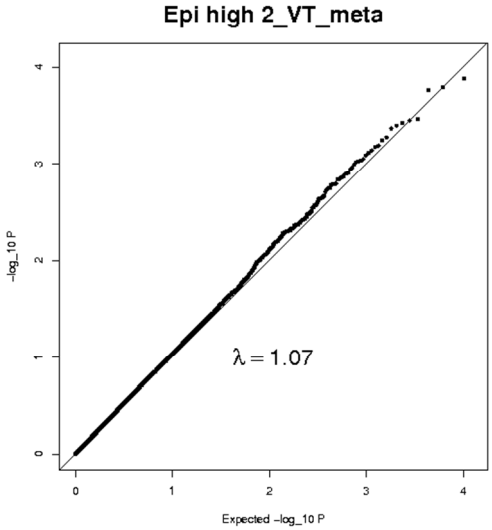


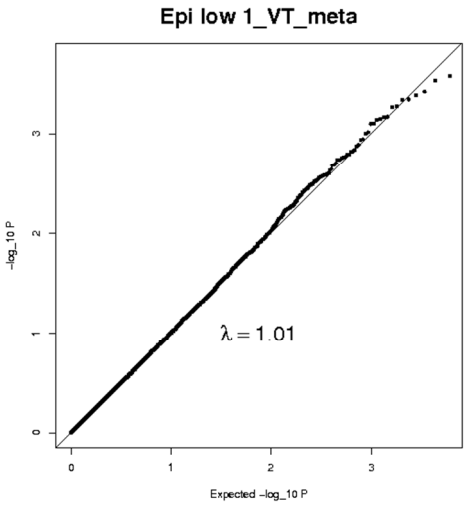


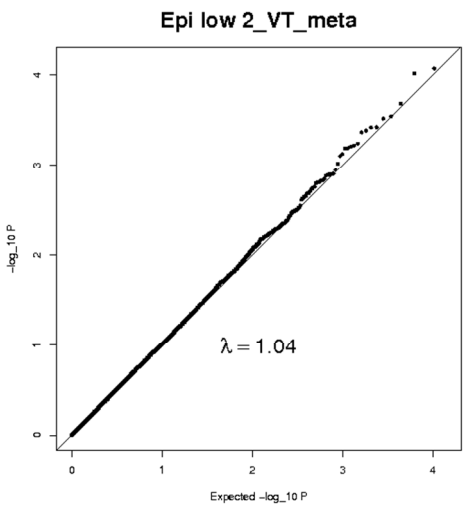

ADP low 1_VT_meta
λ = 1.02
Expected −log_10 P
−log_10 P

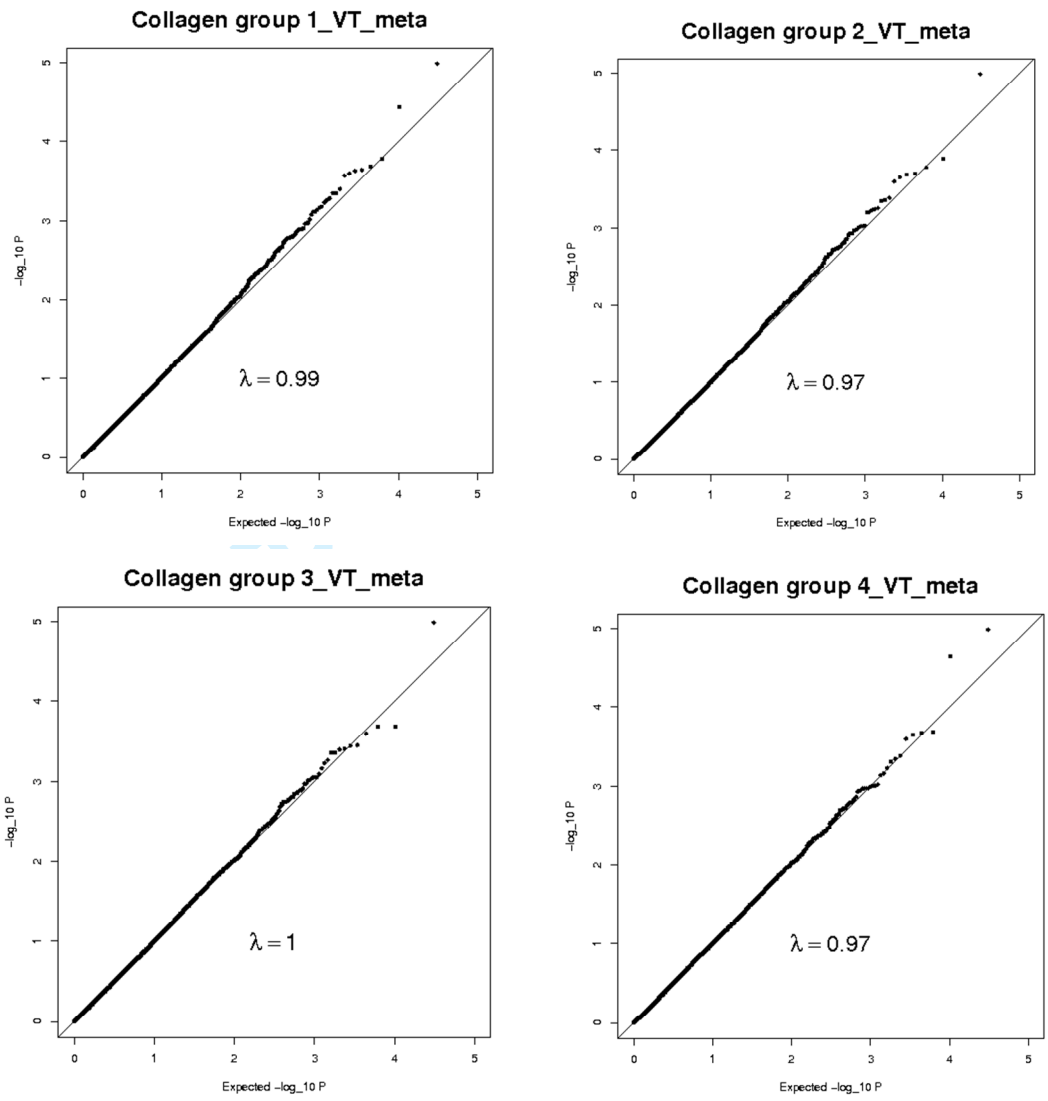

ADP low 2_VT_meta
λ = 0.98
Expected −log_10 P
−log_10 P


Epi high 1_VT_meta
λ = 1.05
Expected −log_10 P
−log_10 P

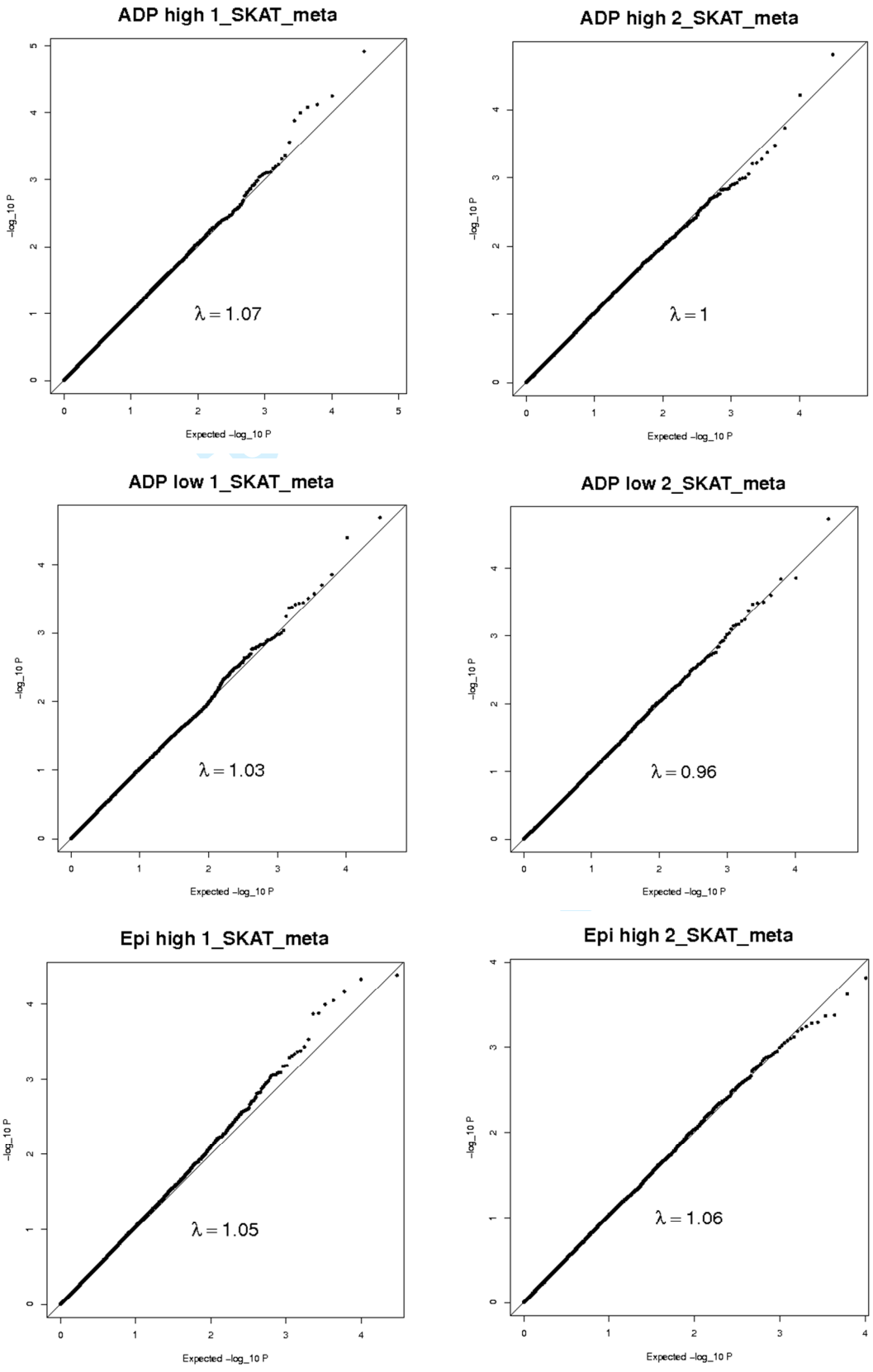

Epi high 2_VT_meta
λ = 1.07
Expected −log_10 P
−log_10 P

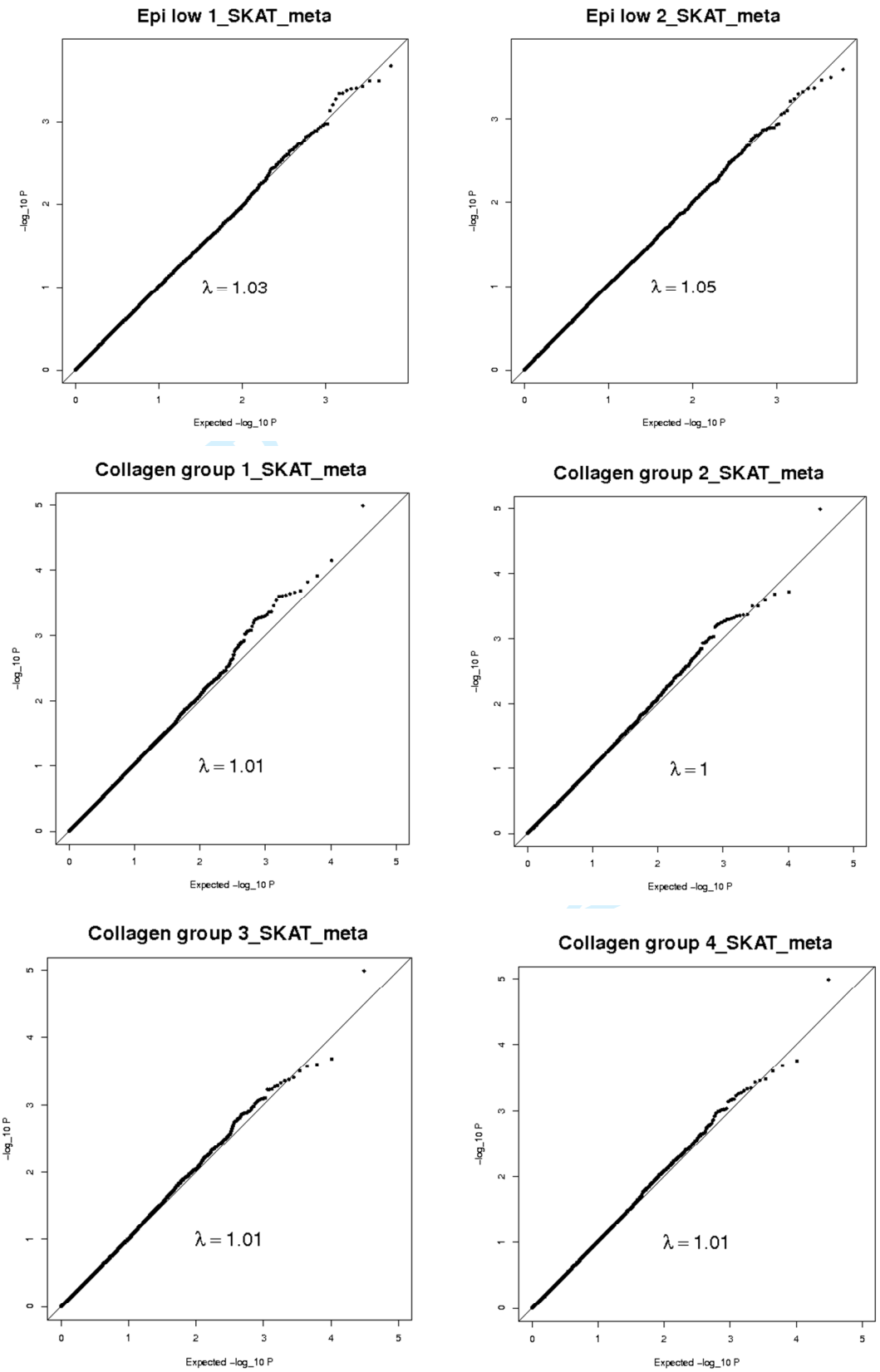

Epi low 1_VT_meta
λ = 1.01
Expected −log_10 P
−log_10 P


Epi low 2_VT_meta
λ = 1.04
Expected −log_10 P
−log_10 P



6. SKAT meta-analysis


ADP high 1_SKAT_meta
λ = 1.07
ADP high 2_SKAT_meta
λ = 1
ADP low 1_SKAT_meta
λ = 1.03
ADP low 2_SKAT_meta
λ = 0.96
Epi high 1_SKAT_meta
λ = 1.05
Epi high 2_SKAT_meta
λ = 1.06
Expected −log_10 P
−log_10 P


Epi low 1_SKAT_meta
λ = 1.03
Epi low 2_SKAT_meta
λ = 1.05
Collagen group 1_SKAT_meta
λ = 1.01
Collagen group 2_SKAT_meta
λ = 1
Collagen group 3_SKAT_meta
λ = 1.01
Collagen group 4_SKAT_meta
λ = 1.01
Expected −log_10 P
−log_10 P