Original Article

OPEN

Associations Between Common and Rare Exonic Genetic Variants and Serum Levels of 20 Cardiovascular-Related Proteins

The Tromsø Study

Terry Solomon, BS; Erin N. Smith, PhD; Hiroko Matsui, MS; Sigrid K. Braekkan, PhD; INVENT Consortium;* Tom Wilsgaard, PhD; Inger Njølstad, MD, PhD; Ellisiv B. Mathiesen, MD, PhD; John-Bjarne Hansen, MD, PhD; Kelly A. Frazer, PhD

Background—Genetic variation can be used to study causal relationships between biomarkers and diseases. Here, we identify new common and rare genetic variants associated with cardiovascular-related protein levels (protein quantitative trait loci [pQTLs]). We functionally annotate these pQTLs, predict and experimentally confirm a novel molecular interaction, and determine which pQTLs are associated with diseases and physiological phenotypes.

Methods and Results—As part of a larger case–control study of venous thromboembolism, serum levels of 51 proteins implicated in cardiovascular diseases were measured in 330 individuals from the Tromsø Study. Exonic genetic variation near each protein's respective gene (*cis*) was identified using sequencing and arrays. Using single site and gene-based tests, we identified 27 genetic associations between pQTLs and the serum levels of 20 proteins: 14 associated with common variation in *cis*, of which 6 are novel (ie, not previously reported); 7 associations with rare variants in *cis*, of which 4 are novel; and 6 associations in *trans*. Of the 20 proteins, 15 were associated with single sites and 7 with rare variants. *cis*-pQTLs for kallikrein and F12 also show *trans* associations for proteins (uPAR, kininogen) known to be cleaved by kallikrein and with NTproBNP. We experimentally demonstrate that kallikrein can cleave proBNP (NTproBNP precursor) in vitro. Nine of the pQTLs have previously identified associations with 17 disease and physiological phenotypes.

Conclusions—We have identified *cis* and *trans* genetic variation associated with the serum levels of 20 proteins and utilized these pQTLs to study molecular mechanisms underlying disease and physiological phenotypes. **(*Circ Cardiovasc Genet*. 2016;9:375-383. DOI: 10.1161/CIRCGENETICS.115.001327.)**

Key Words: biomarker ■ coronary artery disease ■ exome ■ human ■ protein ■ venous thromboembolism

Recent advances in genetics have yielded an unprecedented number of loci associated with disease and are beginning to yield mechanistic insight, such as with the IRX3/5 association with body mass index, which revealed brown adipose as an important regulator of body weight.[1] Genetic variation underlying molecular phenotypes, such as proteins and transcript expression levels, can be important tools in constructing the effects of genetic variations into pathways, ultimately resulting in physiological understanding of diseases.[2] Protein levels, in particular, may be more informative for understanding disease because there is often a poor correlation between transcript and protein levels.[3] Several previous studies[4–6] have systematically identified genetic variations associated with protein levels and isoforms (protein quantitative trait loci [pQTLs]). Although most studies have focused on common variation (minor allele frequency ≥5%), rare variants, which can show strong loss of function effects, can be useful in understanding causality and pinpointing drug targets, such as deletion mutations in *PSCK9* that abolish the PSCK9 protein and reduce low-density lipoprotein cholesterol levels.[7] Systematic screening for rare variation influencing a wide variety of proteins, however, has not yet been performed.

Editorial, see p 318

Clinical Perspective on p 383

Genetic variation is also useful in identifying causal relationship between biomarkers and diseases using tools such as

Received December 16, 2015; accepted June 16, 2016.

From the Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla (T.S.), Department of Pediatrics, Rady's Children's Hospital, San Diego, La Jolla, CA (E.N.S., H.M., K.A.F.); Institute for Genomic Medicine, University of California, San Diego, La Jolla (K.A.F.); Department of Clinical Medicine, K.G. Jebsen Thrombosis Research and Expertise Centre (TREC) (E.N.S., S.K.B., I.N., E.B.M., J.-B.H., K.A.F.), Department of Community Medicine (T.W., I.N.), and Brain and Circulation Research Group, Department of Clinical Medicine (E.B.M.), UiT The Arctic University of Norway; and Division of Internal Medicine, University Hospital of North Norway, Tromsø (S.K.B., J.-B.H.).

*A list of all members in INVENT Consortium is given in the Appendix.

The Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.115.001327/-/DC1.

Correspondence to Kelly A. Frazer, PhD, Frazer Lab, Institute for Genomic Medicine, University of California, San Diego, 9500 Gilman Dr, No. 0761, La Jolla, CA 92093. E-mail kafrazer@ucsd.edu

Mendelian randomization[8] and could be used to ascertain how risk factors differentially affect various diseases, as well as trace causal pathways between risk loci and disease. We are investigating risk factors for cardiovascular diseases, including myocardial infarction (MI) and venous thromboembolism (VTE) in the Tromsø Study,[9] a longitudinal prospective cohort study. We previously assayed 51 cardiovascular-related proteins in 419 first-ever MI cases and 398 controls in serum collected years before the MI event.[10] Of the proteins measured, 17 were predictors for MI when considered individually after adjusting for traditional risk factors. Genetic variation associated with these protein levels could be used to study underlying mechanisms of cardiovascular diseases.

Here, using whole exome sequencing data and HumanCoreExome BeadChips, we investigate whether genetic variants are associated with the serum levels of the same 51 cardiovascular-related proteins in 330 individuals chosen from the Tromsø Study because they did or did not go on to develop VTE during the 18 years of follow-up (mean time to VTE of 9 years). The serum samples were collected at study entry enabling us to identify pQTLs associated with baseline protein levels. We perform both common and rare variation association analyses to identify *cis*-pQTLs. Further characterization of the *cis*-pQTLs to determine whether they also act as *trans*-pQTLs with any of the other 51 cardiovascular-related proteins, recapitulated well-established physiological relationships between F12, kallikrein, uPAR, kininogen, and a recent genetic association with NTproBNP. We experimentally confirmed an inferred physiological interaction from the *trans*-pQTLs by showing that kallikrein cleaves proBNP in vitro. We then examine genetic associations from genome-wide association studies on coronary artery disease (CAD) and VTE, as well as published literature to identify physiological and disease associations.

Methods

The Tromsø Study

The Tromsø Study is a prospective, single-site, cohort study of the inhabitants of Tromsø, Norway. In 1994 to 1995, 27 158 individuals filled out epidemiological surveys and donated (nonfasting) blood to the National CONOR Biobank.[9] These individuals were followed until 2013, with repeated surveys and identified in national registries that report various diseases and causes of death. In 2013, we identified individuals who, between 1995 and 2013, had had an incident of VTE or death due to VTE, regardless of other comorbidities. We chose age- and sex-matched controls randomly from the cohort. These samples were chosen for a currently ongoing case–control study of VTE. DNA and protein levels were ascertained from the blood collected in 1994.

For this specific study, blood and nonfasting serum samples were collected from 330 healthy individuals (166 males and 164 females) aged 45 to 75 years (Table I in the Data Supplement). There were 196 individuals diagnosed with VTE between the study entry (1994–1995) and the 18-year follow-up period (2013) and 134 controls without development of VTE during this period. Aspirin usage and other medication information were not collected for the Tromsø study. DNA was isolated from the blood for genotyping, and serum samples were used to assay protein levels. The regional committee for medical and health research ethics in North Norway approved the study, and all participants gave informed written consent.

Protein Quantification

Protein levels were quantified using the same methods and at the same time as our previous MI study,[10] but the samples from people who went on to develop VTE were not included in that study. Briefly, the literature was searched to create a list of >900 cardiovascular-related proteins that might be potential biomarkers for MI and atherosclerosis. This list was then prioritized to 165 candidate proteins, of which 51 had sufficient commercially available reagents (2 antibodies and purified protein for control) in order for Tethys Bioscience, Inc (Emeryville, CA) to perform successful sandwich enzyme-linked immunosorbent assays (Methods section and Table II in the Data Supplement). All protein levels were quantile normalized and mapped to the normal distribution using qnorm in R, and significance was tested using *Z* scores.

Variant Identification and Annotation

Genotypes were determined using exome sequencing (n=243) or exome genotyping arrays (n=87). Sequences were mapped and called using the Burrows-Wheeler Aligner[11] and Genome Analysis Toolkit,[12] imputed to the 1000 Genomes Project[13] using Beagle,[14] and functionally annotated for predicted effect and regulatory regions (Methods section in the Data Supplement).

Statistical Analysis

Associations were performed using the Efficient and Parallelizable Association Container Toolbox.[15] We used sex, age at study entry, body mass index at study entry, genotyping platform, and VTE case–control status as covariates. Three covariates (age at serum collection, sex and body mass index at serum collection) were associated, respectively, with 10, 10, and 13 of the phenotypes (the 51 protein serum levels) when performing linear regressions, defined as having a false-discovery rate-adjusted P value of <0.05, and they were included for consistency.

For common variants (MAF≥1%), we used the Efficient Mixed Model Association eXpedited[16] (a mixed model implemented in the Efficient and Parallelizable Association Container Toolbox[15]), using q.emmax to test for single-site association. For *cis* associations, we included any imputed common variants located within the interval surrounding and including the gene (±500 kb from transcript start and stop positions) that encodes the protein(s) being tested (C3 and C3b share the same locus). For *cis*-acting-in-*trans* associations, we tested all significantly associated common *cis* variants against each of the other 50 phenotypes. For *trans* associations, we tested the 100 378 common variants found in the 50 intervals against each of the 51 phenotypes (Figure 1).

The optimal Sequence Kernel Association Test[17] was used to test clusters of rare variants (MAF≤5%) for association as implemented in the Efficient and Parallelizable Association Container Toolbox, using the skat-o version of the mmskat test. Rare variants were classified in 3 ways: (1) MAF≤5%, all rare variants located within the gene body and 2kb upstream; (2) deleterious, all rare variants located in the gene body and the 2-kb upstream region that were annotated as stop-gain, stop-loss, start-loss, essential splice site disruption, frame-shift causing, or nonsynonymous using Variant Effect Predictor annotations; and (3) Combined Annotation Dependent Depletion (CADD)-score, all rare variants in the gene or the 2-kb upstream region with a Phred-scaled c-score of >10, as determined by Kircher et al.[18]

We corrected for multiple testing by permuting the phenotype–genotype relationship 1000× and for each permutation performing all variant-phenotype tests for each association type separately (eg, *cis*, *cis*-acting-in-*trans*, or *trans*).[19] We obtained the lowest P value from each permutation across all phenotypes and created a null distribution of minimum P values. An association was considered significant (family-wise P<0.05) if the nominal P value was <95% of the null distribution (Table III in the Data Supplement).

To test for multiple, independent variants in the same locus, the top variant was included as a covariate until there was no longer a significant association (family-wise P<0.05) detected for that protein.

Power Calculations

We calculated power using an equation from the Abecasis laboratory (http://genome.sph.umich.edu/wiki/Power_Calculations:_Quantitative_Traits) for common variants and the Sequence Kernel Association Test R package[20] for rare variants. We had 80% power to detect effects (R^2)

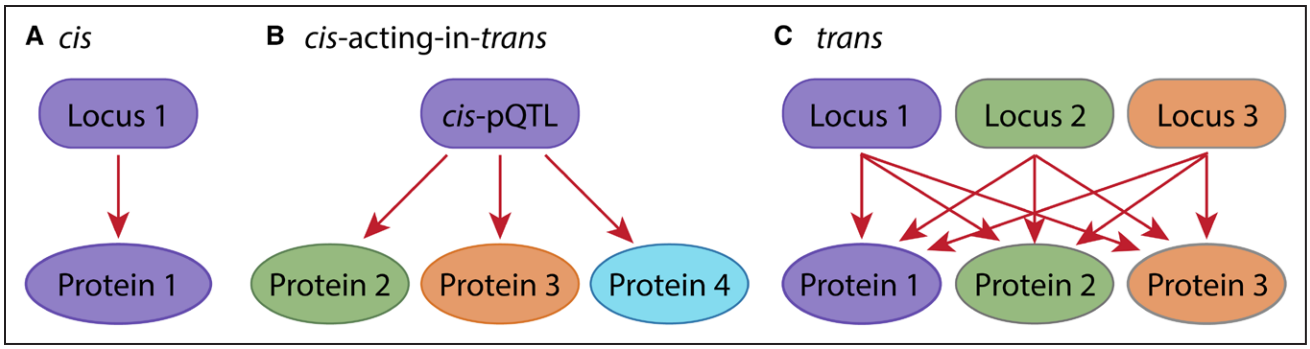


Figure 1. Overview of the 3 stages of association analyses. **A**, *cis*: for each of the 51 phenotypes (protein levels), we tested the variants located in the *cis* gene loci for associations with their respective protein level, (**B**) *cis*-acting-in-*trans*: we tested the significant *cis*-protein quantitative trait loci (pQTLs) from stage 1 for *trans* effects against each of the 50 other protein levels, and (**C**) *trans*: we tested all variants in the 50 *cis* loci (C3 and C3b share the same locus) for association with each of the 51 protein levels.

down to 0.113 for the *cis*, common-variant analysis and effects (β) of 1.25 for the *cis*, rare-variant analysis (assuming that 50% of the variants are causal), which is comparable to other pQTL studies.[4–6,21,22] Further details are in the Data Supplement (Methods section, Figures I and II, and Table IV in the Data Supplement).

Clinical and Molecular Phenotype Association

Significant pQTLs from this study were queried against the expression Quantitative Trait Loci (eQTLs) found by Schadt et al[23] in liver cells and the Gene-Tissue Expression database[24] (version 4, build 200, accessed at http://www.gtexportal.org/home/) for all tissue types. In addition, we determined whether they (or a variant in linkage disequilibrium [LD]) overlapped any of the variants identified as pQTLs in 5 similar-sized independent studies that investigated protein levels in serum[4] or plasma.[4–6,22,25] We examined pQTLs for clinical significance by determining whether the variant has been previously identified and submitted to the Online Mendelian Inheritance in Man,[26] the genome-wide association study (GWAS) Catalog,[27] or the Genome-Wide Repository of Associations between Single nucleotide polymorphisms and Phenotypes v2.0.[28] We identified pQTLs that were also significant in large meta-analyses of individuals of European descent for CAD or VTE. Data on CAD was downloaded from http://www.cardiogramplusc4d.org. For this analysis, we only used the results from the Coronary Artery Disease Genome wide Replication and Meta-analysis (CARDIoGRAM) genome-wide association study[29] because these individuals are of European descent. Data on VTE was shared by the International Network against Thrombosis (INVENT Consortium).[30]

In Vitro Assay of proBNP Cleavage

We obtained native kallikrein from human plasma from EMD-Millipore (Darmstadt, Germany; cat no. 420307); recombinant proBNP from Abcam (Cambridge, Ma; cat no. ab151881); and the kallikrein inhibitor, H-D-Phe-Phe-Arg-chloromethyl ketone (PPACK II), from Santa Cruz Biotechnology (Dallas, TX; cat no. sc-203215). A total of 354 ng (374 nmol/L) of kallikrein was incubated with 80 ng (606 nmol/L) of proBNP with and without 26.5 ng (36.4 µmol/L) of PPACK II for 30, 60, and 90 minutes at 37°C. The reactions were stopped by adding 4× lithium dodecyl sulfate sample buffer and dithiothreitol, and heating them for 2 minutes at 85°C. The proteins were run on a Tricine-SDS–PAGE gel from ThermoFisher (Waltham, MA), and either detected using the SilverQuest Silver Staining Kit from ThermoFisher or transferred to a polyvinylidene difluoride membrane and detected using an anti-BNP antibody from Novus Biologicals (Littleton, Co; cat no. NB100-62133) and chemiluminescence.

Results

Study Overview

The subjects were chosen as a substudy from an ongoing case–control study examining the genetics of VTE, and includes 196 individuals who developed VTE during the 18-year follow-up and 134 individuals who did not (Table I in the Data Supplement). Serum was assayed for the levels of 51 proteins using enzyme-linked immunosorbent assays (Table II in the Data Supplement). On average, we obtained high-quality protein measurements for 311 individuals per phenotype. We investigated whether any of the protein levels were associated with VTE case–control status and found no significant associations. Knowing that the protein levels were not statistically associated with VTE enabled us to combine the VTE cases and controls to explore the effects of genetic variation on baseline protein levels.

We performed high coverage (≈100×) exome sequencing on DNA from blood samples for 243 individuals and assayed an additional 87 with HumanCoreExome Beadchips. We identified 158 137 variants (direct genotyping and imputation) in the 50 intervals that encode the 51 proteins (Table V in the Data Supplement). The majority of imputed variants were intergenic or intronic because these variants were not already captured by the genotyping array or were outside of the exome-sequencing target regions (Table VI in the Data Supplement). There was an average of 1122 variants per locus with the *AGER* locus having the most variants (3523) and the *CD40LG* locus having the fewest (441; Table II in the Data Supplement).

Identifying *cis*-pQTLs From Common Variants

To identify genetic variation associated with serum protein levels, we tested for association between variants within the gene's *cis* locus and the normalized protein level for each of the 51 protein levels, adjusting for sample relatedness and population structure using a kinship matrix and including age, sex, body mass index, genotype platform, and subsequent VTE status as covariates. Because of the high likelihood of linkage disequilibrium at the *cis* loci and slight correlations among protein levels, we accounted for multiple testing by performing permutations to obtain a family-wise error rate. We identified significant associations (adjusted $P<0.05$, nominal $P<6.97\times10^{-7}$; Table 1; Figure 2) for 13 of the 51 phenotypes. To test for multiple, independent associations, we performed sequential conditioning on the most highly associated variant and found 2 independent *cis* associations for LP(a). Of the 14 *cis*-pQTLs that we report, we have replicated

Table 1. Significant *cis*-pQTLs From the Common-Variant Association Analysis

Protein	Gene	Top Variant	Chr (b37)	Position (b37)	Alleles (Ref/Alt)	Alt Allele Frequency	Nominal *P* Value	Adjusted *P* Value	β	R^2
a2-AP	*SERPINF2*	rs8077638	17	1640793	C/T	0.19	5.4×10^{-37}	<0.001	−1.15	0.42
AGT	*AGT*	rs4762	1	230845977	G/A	0.14	4.4×10^{-22}	<0.001	1.08	0.25
ANG	*ANG*	rs3748338	14	21167576	A/T	0.11	1.9×10^{-12}	<0.001	0.86	0.16
C3	*C3*	rs11569415	19	6716279	G/A	0.15	6.9×10^{-10}	<0.001	−0.63	0.13
C3B	*C3*	rs2230199	19	6718387	G/C	0.23	1.2×10^{-12}	<0.001	−0.65	0.16
CHIT1	*CHIT1*	rs2486951	1	203174921	A/G	0.18	3.7×10^{-21}	<0.001	−1.01	0.26
F12	*F12*	rs1801020	5	176836532	A/G	0.76	2.5×10^{-32}	<0.001	0.99	0.38
KLKB1	*KLKB1*	rs3733402	4	187158034	G/A	0.53	4.4×10^{-12}	<0.001	−0.51	0.15
KNG1	*KNG1*	rs166479	3	186443250	T/C	0.41	1.7×10^{-10}	<0.001	−0.46	0.13
LBP	*LBP*	rs2232613	20	36997655	C/T	0.10	2.2×10^{-22}	<0.001	−1.20	0.27
LP(a)*	*APOA*	rs41272114	6	161006077	C/T	0.030	3.1×10^{-8}	0.002	−1.27	0.10
LP(a)*	*APOA*	rs56393506	6	161089307	C/T	0.083	1.7×10^{-7}	0.011	0.66	0.08
MMP3	*MMP3*	rs7926920	11	102698724	G/A	0.35	2.4×10^{-14}	<0.001	−0.41	0.17
MMP8	*MMP8*	rs35231465	11	102584135	G/A	0.036	1.9×10^{-7}	0.012	−1.10	0.09

β indicates effect size of association in standard deviation units per each copy of the alternate allele; Alt, alternate; pQTLs, protein quantitative trait loci; R^2, amount of phenotypic variation explained by the variant; and Ref, reference.

*LP(a) has 2 independent *cis*-pQTLs. rs56393506 was identified as an independent pQTL for LP(a) by performing the association analysis using genotypes from the top variant (rs41272114) as a covariate.

8 known pQTLs and identified 6 novel pQTLs. The same variant or a variant in LD (r^2>0.5 in EUR) has been previously reported for 8 proteins with the same direction of effect that we found: AGT,[22] C3,[5] C3b,[5] CHIT1,[6] F12,[6,25] LBP,[6] one of the variants for LP(a),[31] and MMP3[32] (Table VII in the Data Supplement). Of the 6 novel pQTLs that we identified, 4 proteins have not previously been reported to have a *cis*-pQTL (a2-AP, ANG, KLKB1, and MMP8) and 2 proteins have been previously associated with a pQTL, but the variant identified here is not in LD with the previous variant (KNG1[25,6] and LP(a)[31]). rs3373402 in *KLKB1* was previously reported to affect KLKB1 binding with kininogen (KNG1) but did not affect KLKB1 levels in plasma[33]; therefore, although this variant has been previously functionally characterized, this is a novel pQTL. We annotated the 14 pQTLs for functional effects and identified their chromatin state in the tissue where their target gene is most highly expressed (Table VIII Data Supplement). Ten of the 13 proteins are predominantly secreted by the liver. Five of the top variants are missense variants, 3 are in the untranslated regions, and 5 lie in predicted regulatory regions based on chromatin state annotations. These analyses suggest possible mechanisms of action for some of the *cis*-pQTLs.

Identifying *cis*-pQTLs From Rare Variation

We next tested whether the combination of multiple rare variants at each *cis*-locus was associated with protein levels. There were 3675 rare variants identified across all 50 loci. For rare variation association analyses, rare variants are grouped according to frequency or function and then jointly tested for association. Because functional prediction methods vary and it is currently unknown what method is superior,[34] we used 3 different classifications (MAF, Deleterious, and

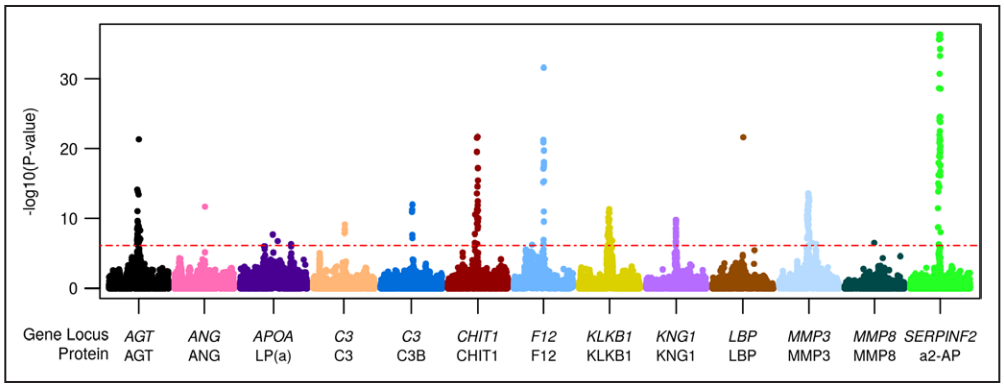


Figure 2. Association of *cis* variants with protein levels. Modified Manhattan plot showing the −log10 *P* values for association between variants in each *cis* locus (interval encoding protein ±500 kb) and the respective protein levels. The red dashed line indicates the study-wide significant *P*-value cutoff when only examining *cis* regions (6.9×10^{-7}) for a family-wise error rate <0.05.

Table 2. Rare Variant *cis*-pQTLs That Are Significant Using At Least 1 of the 3 Grouping Methods

	MAF≤5%		Deleterious		CADD10	
Protein	Nominal *P* Value	Adjusted *P* Value	Nominal *P* Value	Adjusted *P* Value	Nominal *P* Value	Adjusted *P* Value
AGER	3.2×10^{-4}	0.041	0.003	n.s.	0.006	n.s.
CD40L	0.042	n.s.	9.8×10^{-5}	0.009	0.003	n.s.
CHIT1*	4.3×10^{-8}	<0.001	0.108	n.s.	0.127	n.s.
Fetuin A	2.5×10^{-4}	0.026	1.5×10^{-5}	0.002	2.4×10^{-6}	<0.001
LP(a)	1.1×10^{-5}	0.002	2.6×10^{-8}	<0.001	4.4×10^{-7}	<0.001
MMP8	0.247	n.s.	7.7×10^{-4}	n.s.	6.3×10^{-6}	0.002
TAFI	0.014	n.s.	5.2×10^{-5}	0.003	0.002	n.s.
TIMP4	0.050	n.s.	2.4×10^{-4}	0.026	1.7×10^{-4}	0.018

n.s. indicates not significant; and pQTLs, protein quantitative trait loci.
*Not significant after adjusting for the common pQTL (rs2486951).

CADD-score—see Methods section of this article). Across all loci, there was a range of 1 to 90 variants used for each method, with the MAF method having the most rare variants and CADD scores having the fewest. To account for multiple testing, we tested all the 3 classifications in each round of permutations to determine the family-wise error rate *P*-value cutoff. We performed an optimal Sequence Kernel Association Test using the same covariates as for the common-variant association. We identified 8 *cis*-pQTLs that were significant using ≥1 classifications (adjusted $P<0.05$, nominal $P<3.72\times10^{-4}$; Table 2; Table IX in the Data Supplement). Of these, *cis* rare variation has been associated with AGER,[35] Fetuin A,[36] and LP(a) levels[31]; to our knowledge, the other 5 associations are novel.

Of the 8 proteins associated with rare variation, 3 were also associated with a common pQTL (CHIT1, LP(a), and MMP8). For LP(a) and MMP8, a common pQTL (with an MAF<5%) was also present on the list of rare variants and removal of these from the rare-variant analysis made the rare association nonsignificant (CADD nominal *P* value, 0.148 and 0.469, respectively). For CHIT1, the common pQTL had an MAF of 18% and although not on the list of rare variants, when we included this variant as a covariate in the rare-variant analysis the association was nullified (nominal *P* value=0.147). These results suggest that the rare variants in the *CHIT1* locus were associated with CHIT1 serum levels because of linkage disequilibrium with the common pQTL. Because the driving variant was common, we do not consider the CHIT1 association to be valid, resulting in 7 proteins associated with rare variants.

Identifying *trans*-pQTLs

To characterize potential downstream effects of *cis*-pQTLs, we investigated whether any of the common *cis*-pQTLs might also have *trans* effects (*cis*-acting-in-*trans*) on any of the other 50 protein levels. After permutation to obtain adjusted *P* values, we identified 2 *cis*-acting-in-*trans* loci, each of which was significantly associated with 3 proteins (adjusted $P<0.05$, nominal $P<7.29\times10^{-5}$; Table 3). There was significant overlap in the proteins associated with the 2 loci, and the associations were consistent with known physiological relationships between F12, KLKB1, KNG1, and uPAR, and the recently reported genetic relationship with NTproBNP[37] (Figure 3), despite none of the protein levels being strongly correlated (Figure 3; Table X in the Data Supplement). We did not observe an association between the *cis*-pQTL for *KLKB1* and F12 protein levels, despite the known physiological relationships of KLKB1 and F12 (Figure 3). Importantly, the genetic associations of *KLKB1* and *F12* with NTproBNP suggest that KLKB1 may physiologically cleave proBNP (the NTproBNP precursor). These findings illustrate how genetic variation can be used to identify potentially novel physiological relationships among proteins.

We further performed a full pairwise association (*trans*) between any of the variants located in the 50 regions encoding the proteins used in this study and all 51 protein levels. After permutation adjusting (adjusted $P<0.05$, nominal $P<1.25\times10^{-8}$), we did not find any additional *trans* associations and none of the *cis*-acting-in-*trans* associations remained significant; however, 11 of the 14 *cis* associations remained significant.

Using a similar approach to the common variants, we tested if any of the rare variant *cis*-pQTLs were associated with any of the other 50 protein levels and did not observe any significant associations (adjusted $P<0.05$, nominal $P<5.30\times10^{-5}$). In addition, we tested all 50 *cis* regions against all 51 protein levels in a pairwise manner, but did not identify additional associations (adjusted $P<0.05$, nominal $P<9.21\times10^{-6}$), although 4 of the 8 rare *cis* associations were still significant at the more stringent threshold.

Role of Kallikrein in proBNP Maturation

We experimentally tested the *cis*-acting-in-*trans* associations suggesting that kallikrein (KLKB1) may physiologically cleave proBNP. ProBNP is produced as a propeptide that may be cleaved intracellularly into BNP and NTproBNP, 2 biomarkers for heart failure,[38] before being secreted by cardiomyocytes in response to cardiac stress. Intracellularly, it is thought that furin or corin cleave proBNP,[39] but it is unclear which enzyme cleaves proBNP extracellularly when it is secreted intact.[40] To test whether kallikrein can cleave proBNP in vitro, we incubated increasing concentrations of kallikrein

Table 3. *cis*-pQTLs That Also Act As *trans*-pQTLs

Variant	Protein	Nominal *P* Value	Adjusted *P* Value	β	R^2
rs1801020 in the *F12* locus	F12	2.5×10^{-32}	<0.001	0.985	0.382
	KLKB1	5.4×10^{-8}	<0.001	−0.488	0.092
	KNG1	1.1×10^{-7}	<0.001	−0.479	0.097
	NTproBNP	1.2×10^{-5}	0.002	−0.380	0.061
rs3733402 in the *KLKB1* locus	KLKB1	4.4×10^{-12}	<0.001	−0.506	0.152
	KNG1	5.2×10^{-5}	0.034	−0.309	0.049
	NTproBNP	4.2×10^{-8}	<0.001	−0.393	0.098
	uPAR	4.4×10^{-8}	<0.001	−0.401	0.097

β indicates effect size of association in standard deviation units; pQTLs, protein quantitative trait loci; and R^2, amount of phenotypic variation explained by the variant.

(74.8, 374, 748, and 1497 nmol/L) with proBNP for 1 hour at body temperature (37°C) and saw progressive depletion of proBNP levels (Figure IV in the Data Supplement). This depletion was prevented with the addition of PPACK II, a kallikrein-specific inhibitor. From this, we chose to incubate 374 nmol/L of kallikrein with proBNP for 30, 60, or 90 minutes and again, we saw that the levels of proBNP decreased (Figure 4). These results suggest that kallikrein has the ability to cleave proBNP in vivo.

Annotation of pQTLs Using Existing Databases and GWAS

We investigated whether the 14 common pQTLs that we identified were reported as eQTLs in the Gene-Tissue Expression database (ref 24) or identified in the liver eQTL dataset from Schadt et al.[23] In the Gene-Tissue Expression database, the AGT pQTL was identified as an eQTL in 10 tissues (*P* values from of 2.0×10^{-6} to 1.3×10^{-33}), the CHIT1 pQTL is an eQTL in whole blood (*P* value 4.2×10^{-8}), the F12 pQTL is an eQTL in liver (*P* value 2.3×10^{-10}), and the pQTL in the *SERPINF2* locus (a2-AP protein) is an eQTL in 6 tissues (*P* values from 5.3×10^{-7} to 8.8×10^{-18}). In addition, the pQTLs for a2-AP, AGT, CHIT1, F12, KLKB1, and MMP3 were also identified as eQTLs for other nearby genes. In the Schadt et al's data set rs3748338 in the *ANG* locus is in LD (r^2=0.24) with an eQTL for ANG (rs8008440). Thus, of the 14 common pQTLs, 2 have previously been identified as an eQTL for the *cis* gene, 3 as an eQTL for both the *cis* gene and other nearby genes, and 3 as an eQTL for nearby gene(s).

We also looked up whether there are any known disease associations with the 14 pQTLs that we identified using the GWAS catalog.[27] the Genome-Wide Repository of Associations between Single nucleotide polymorphisms and Phenotypes,[28] and the Online Mendelian Inheritance in Man[26] (Table VII in the Data Supplement). The 8 known pQTLs along with the kallikrein pQTL are associated with a variety of phenotypes, including age-related macular degeneration (C3b), activated partial thromboplastin times (F12), serum metabolites (KLKB1), binding of LBP to lipopolysaccharide (LBP), and plasma plasminogen levels (LP(a)). In total, 9 pQTLs have been associated with 17 disease or physiological phenotypes.

Finally, to investigate whether the pQTLs identified here are associated with VTE or CAD, we examined the results of 2 previously published meta-analyses. The INVENT[30] study is a large meta-analysis of 7507 cases and 52 632 controls to identify variants associated with VTE. The CARDIoGRAM[29] study is a large meta-analysis of 22 233 cases and 64 762 controls designed to identify variants associated with CAD, which is predominantly composed of MI. Of 14 common pQTLs, 10 (71.4%) could be tested in the INVENT and CARDIoGRAM datasets (Table XI in the Data Supplement). The KLKB1 pQTL (rs3733402) is significantly associated with VTE; however, this association becomes nonsignificant when the analysis is conditioned on the top 6 SNPs associated with VTE from the literature. The KLKB1 pQTL (rs3733402) is also nominally associated with CAD (*P*=0.0086). The KNG1 pQTL (rs166479) had a nominal *P* value of <0.05 in the INVENT consortium. Although one of the pQTLs for LP(a) (rs41272114) has previously been associated with CAD,[31] it was not present in either data set. In addition, among the 17 protein biomarkers that we previously identified as being associated with the first MI,[10] we identified common *cis*-pQTLs for 6 (C3, C3b, KLKB1, LP(a), MMP3, and MMP8) and rare *cis*-pQTLs for 5 (LP(a), MMP8, TAFI, and TIMP4). Although we found pQTLs for these MI biomarkers, they were not associated with CAD in the CARDIoGRAM study, which could indicate that the biomarkers are not causally related to CAD, but may be a result of the relatively small sample size in the GWAS compared with typical Mendelian

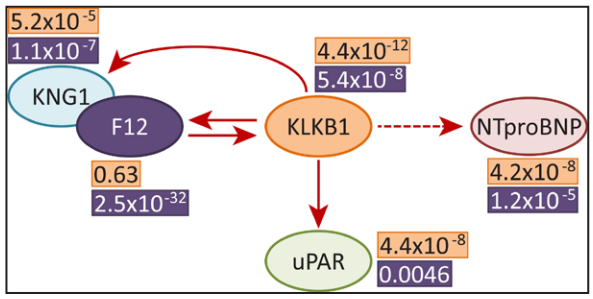


Figure 3. Schematic diagram showing proteins with identified *trans* associations and their nominal associations with variants in *F12* and *KLKB1*. Previously known (solid) and proposed in this study (dashed) cleavage reactions are represented with arrows. Nominal *P* values for the associations between protein levels and rs3733402 in the *KLKB1* locus and rs1801020 in the *F12* locus are shown, respectively, in orange and purple boxes next to the protein of interest.

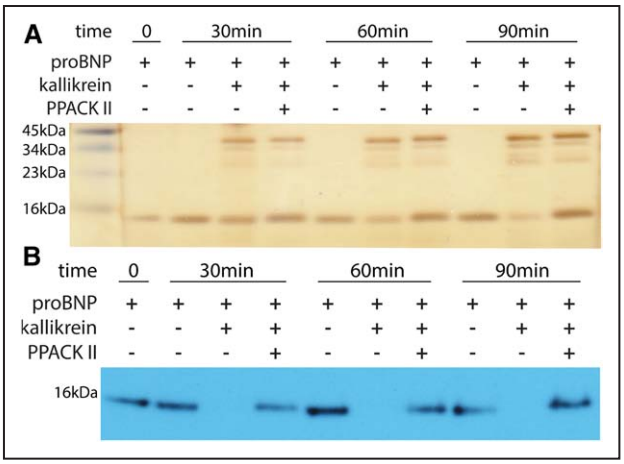


Figure 4. Kallikrein cleaves proBNP in vitro. **A**, A silver stain of recombinant proBNP and kallikrein incubated together for 30, 60, and 90 min with and without a kallikrein-specific inhibitor (PPACK II) and (**B**) a western blot of an identical experimental setup using an anti-BNP antibody. The silver stain binds all protein present and is a more sensitive procedure than using the anti-BNP antibody for the Western blot. We think that this explains why the amount of proBNP in the +/+/– wells visually seems to be different between the silver stain and Western blot.

randomization studies. Thus, although CAD and VTE were not significantly associated with pQTLs, these loci could be used in further larger studies to elucidate functional mechanisms underlying disease.

Discussion

Using a combination of exome sequencing and exome arrays in 330 individuals, we identified 27 genetic associations between pQTLs and the serum levels of 20 proteins: 14 associations with common variation in *cis*, of which 6 are novel and have not been previously reported; 7 associations with rare variants in *cis*, of which 4 are novel; and 6 associations in *trans*. Ultimately, 15 proteins were associated with single sites and 5 were associated with rare variants. The strongest associations were identified for *cis* variation near the gene locus, but by directly testing the *cis*-pQTLs, we also identified 2 that acted in *trans*. Despite the limitations of our study (including a relatively small sample size and lack of a formal replication cohort), the presence of robust associations suggest that exome analysis is an effective tool to identify genetic variation associated with serum protein levels and that larger sample sizes would likely capture additional *trans* effects.

This is the first study, to the best of our knowledge, that uses exome data to investigate the effects of both common and rare variation on more than 50 protein levels; and thus, it provides insight into rare variant association methods. For rare-variant analysis, we used 3 different methods for grouping variants within a gene and accounted for the additional testing through permutation. Some associations were consistent across all the 3 methods, such as LP(a), which carried a large number of variants (Table IX in the Data Supplement) and for which rare variation has previously been associated with the protein level in the blood.[41] Others were only significant in 1 test, such as MMP8 when variants were grouped based on CADD score, which could be because of few variants with weak effects and would benefit from larger sample sizes to include more predicted functional sites. Variants with an MAF between 1% and 5% were tested in both the common- and rare-variant analyses. In 2 cases (LP(a) and MMP8), adjusting for the top common pQTL (with an MAF<5%) nullified the association. In addition, for CHIT1, common variation (MAF>5%) was associated with rare variants through cryptic LD and adjusting for the common variant also nullified the association. These data suggest that significant common and rare single sites may drive gene-based rare variant associations.

Of the 14 common pQTLs, 4 are missense variants in the relevant gene. Of the 10 other variants, 3 are intronic, 2 are in the exons of nearby genes, and 5 lie in regions that are predicted to have regulatory functions, such as interrupting protein-binding sites or splicing (Table VIII in the Data Supplement). Analysis of the function of sequences harboring the pQTL can elucidate the mechanism of the variant. For example, it has been shown that rs1801020 in the 3′ untranslated region of the *F12* locus prevents translation of F12.[42] The mechanisms of the other 4 regulatory pQTLs are not yet understood, but the results shown here point to plausible mechanisms. For instance, ANG and RNASE4 are isoforms of the same gene with different functions and differential expression patterns that are influenced by CTCF.[43] The ANG pQTL is in the last exon of *RNASE4*, near a CTCF-binding site that affects isoform expression levels.[43] This, and other potentially regulatory pQTLs, could be functionally tested using in vitro and in vivo assays for changes in gene or isoform expression. Thus, although we focused on exome sequences to generate genotypes for this analysis, imputation enabled us to identify many pQTLs with predicted regulatory effects.

pQTLs can be used to understand the relationship between proteins and disease, either through tracing molecular impacts through pathways or through studies of Mendelian randomization. By examining potential *trans* associations with *cis*-pQTLs, we recapitulated known and recently reported relationships between these proteins. The relationships between F12, kallikrein, and kininogen comprise the start of the intrinsic coagulation pathway,[44] the association between kallikrein and uPAR has been previously explored,[45] and the genetic relationship between kallikrein and NTproBNP was identified in a recent GWAS.[37] We show that kallikrein is able to cleave proBNP in vitro using purified reagents, suggesting that extracellularly, kallikrein could be responsible for cleaving proBNP into NTproBNP and BNP, although further experiments are necessary to verify that this reaction occurs naturally in plasma. We also identified 17 reported disease and physiological phenotype associations with 9 of the pQTLs (8 previously known and 1 novel). Interestingly, 5 of the 6 novel pQTLs were not implicated in GWAS studies. This could reflect a bias in GWAS phenotypes studied or candidate proteins chosen for pQTL studies and supports further work identifying downstream effects of these loci. We observed a nominal association between KLKB1 and CAD, which we previously identified as a biomarker for MI, supporting further examination of this relationship in larger studies. Overall, these findings support the use of pQTLs to identify molecular

and phenotypic effects of proteins and help to elucidate underlying mechanisms of disease.

Appendix

The INVENT Consortium is comprised Philippe Amouyel, Mariza de Andrade, Saonli Basu, Claudine Berr, Jennifer A. Brody, Daniel I. Chasman, Jean-Francois Dartigues, Aaron R. Folsom, Marine Germain, Hugoline de Haan, John Heit, Jeanine Houwing-Duitermaat, Christopher Kabrhel, Peter Kraft, Grégoire Legal, Sara Lindström, Ramin Monajemi, Pierre-Emmanuel Morange, Bruce M. Psaty, Pieter H. Reitsma, Paul M. Ridker, Lynda M. Rose, Frits R. Rosendaal, Noémie Saut, Eline Slagboom, David Smadja, Nicholas L. Smith, Pierre Suchon, Weihong Tang, Kent D. Taylor, David-Alexandre Trégouët, Christophe Tzourio, Marieke C.H. de Visser, Astrid van Hylckama Vlieg, Lu-Chen Weng, and Kerri L. Wiggins.

Sources of Funding

This work was supported by an independent grant from the K.G. Jebsen Foundation in Norway and partially funded by Tethys Bioscience. T. Solomon is supported by an institutional award to the UCSD Genetics Training Program from the National Institute for General Medical Sciences, T32 GM008666.

Disclosures

None.

References

1. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med*. 2015;373:895–907. doi: 10.1056/NEJMoa1502214.
2. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature*. 2009;461:218–223. doi: 10.1038/nature08454.
3. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*. 1997;18:533–537. doi: 10.1002/elps.1150180333.
4. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet*. 2008;4:e1000072. doi: 10.1371/journal.pgen.1000072.
5. Johansson Å, Enroth S, Palmblad M, Deelder AM, Bergquist J, Gyllensten U. Identification of genetic variants influencing the human plasma proteome. *Proc Natl Acad Sci U S A*. 2013;110:4673–4678. doi: 10.1073/pnas.1217238110.
6. Lourdusamy A, Newhouse S, Lunnon K, Proitsi P, Powell J, Hodges A, et al; AddNeuroMed Consortium; Alzheimer's Disease Neuroimaging Initiative. Identification of cis-regulatory variation influencing protein abundance levels in human plasma. *Hum Mol Genet*. 2012;21:3719–3726. doi: 10.1093/hmg/dds186.
7. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med*. 2006;354:1264–1272. doi: 10.1056/NEJMoa054013.
8. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27:1133–1163. doi: 10.1002/sim.3034.
9. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njølstad I. Cohort profile: the Tromso Study. *Int J Epidemiol*. 2012;41:961–967. doi: 10.1093/ije/dyr049.
10. Wilsgaard T, Mathiesen EB, Patwardhan A, Rowe MW, Schirmer H, Løchen ML, et al. Clinically significant novel biomarkers for prediction of first ever myocardial infarction: the Tromsø Study. *Circ Cardiovasc Genet*. 2015;8:363–371. doi: 10.1161/CIRCGENETICS.113.000630.
11. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760. doi: 10.1093/bioinformatics/btp324.
12. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;11:11.10.1–11.10.33.
13. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.
14. Browning BL, Browning SR. Genotype imputation with millions of reference samples. *Am J Hum Genet*. 2016;98:116–126. doi: 10.1016/j.ajhg.2015.11.020.
15. EPACTS [Computer Program]. Version 3.2.5: University of Michigan Center for Statistical Genetics; 2014.
16. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. 2010;42:348–354. doi: 10.1038/ng.548.
17. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*. 2012;13:762–775. doi: 10.1093/biostatistics/kxs014.
18. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310–315. doi: 10.1038/ng.2892.
19. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005;6:95–108. doi: 10.1038/nrg1521.
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. doi: 10.1016/j.ajhg.2011.05.029.
21. Garge N, Pan H, Rowland MD, Cargile BJ, Zhang X, Cooley PC, et al. Identification of quantitative trait loci underlying proteome variation in human lymphoblastoid cells. *Mol Cell Proteomics*. 2010;9:1383–1399. doi: 10.1074/mcp.M900378-MCP200.
22. Kim S, Swaminathan S, Inlow M, Risacher SL, Nho K, Shen L, et al; Alzheimer's Disease Neuroimaging Initiative (ADNI). Influence of genetic variation on plasma protein levels in older adults using a multi-analyte panel. *PLoS One*. 2013;8:e70269. doi: 10.1371/journal.pone.0070269.
23. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol*. 2008;6:e107. doi: 10.1371/journal.pbio.0060107.
24. Consortium G. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580–585.
25. Liu Y, Buil A, Collins BC, Gillet LC, Blum LC, Cheng LY, et al. Quantitative variability of 342 plasma proteins in a human twin population. *Mol Syst Biol*. 2015;11:786.
26. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*. 2005;33(Database issue):D514–D517. doi: 10.1093/nar/gki033.
27. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(Database issue):D1001–D1006. doi: 10.1093/nar/gkt1229.
28. Eicher JD, Landowski C, Stackhouse B, Sloan A, Chen W, Jensen N, et al. GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res*. 2015;43(Database issue):D799–D804. doi: 10.1093/nar/gku1202.
29. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al; Cardiogenics; CARDIoGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333–338. doi: 10.1038/ng.784.
30. Germain M, Chasman DI, de Haan H, Tang W, Lindström S, Weng LC, et al; Cardiogenics Consortium. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am J Hum Genet*. 2015;96:532–542. doi: 10.1016/j.ajhg.2015.01.019.
31. Kyriakou T, Seedorf U, Goel A, Hopewell JC, Clarke R, Watkins H, et al; PROCARDIS Consortium. A common LPA null allele associates with lower lipoprotein(a) levels and coronary artery disease risk. *Arterioscler Thromb Vasc Biol*. 2014;34:2095–2099. doi: 10.1161/ATVBAHA.114.303462.
32. Zhu C, Odeberg J, Hamsten A, Eriksson P. Allele-specific MMP-3 transcription under *in vivo* conditions. *Biochem Biophys Res Commun*. 2006;348:1150–1156. doi: 10.1016/j.bbrc.2006.07.174.
33. Katsuda I, Maruyama F, Ezaki K, Sawamura T, Ichihara Y. A new type of plasma prekallikrein deficiency associated with homozygosity for

Gly104Arg and Asn124Ser in apple domain 2 of the heavy-chain region. *Eur J Haematol*. 2007;79:59–68. doi: 10.1111/j.1600-0609.2007.00871.x.
34. Santorico SA, Hendricks AE. Progress in methods for rare variant association. *BMC Genet*. 2016;17(suppl 2):6. doi: 10.1186/s12863-015-0316-7.
35. Hudson BI, Carter AM, Harja E, Kalea AZ, Arriero M, Yang H, et al. Identification, classification, and expression of RAGE gene splice variants. *FASEB J*. 2008;22:1572–1580. doi: 10.1096/fj.07-9909com.
36. Yuasa I, Umetsu K. Genetic polymorphism of human alpha 2HS-glycoprotein: characterization and application to forensic hemogenetics. *Electrophoresis*. 1988;9:404–410. doi: 10.1002/elps.1150090810.
37. Musani SK, Fox ER, Kraja A, Bidulescu A, Lieb W, Lin H, et al. Genome-wide association analysis of plasma B-type natriuretic peptide in blacks: the Jackson Heart Study. *Circ Cardiovasc Genet*. 2015;8:122–130. doi: 10.1161/CIRCGENETICS.114.000900.
38. Clerico A, Fontana M, Zyw L, Passino C, Emdin M. Comparison of the diagnostic accuracy of brain natriuretic peptide (BNP) and the N-terminal part of the propeptide of BNP immunoassays in chronic and acute heart failure: a systematic review. *Clin Chem*. 2007;53:813–822. doi: 10.1373/clinchem.2006.075713.
39. Semenov AG, Tamm NN, Seferian KR, Postnikov AB, Karpova NS, Serebryanaya DV, et al. Processing of pro-B-type natriuretic peptide: furin and corin as candidate convertases. *Clin Chem*. 2010;56:1166–1176. doi: 10.1373/clinchem.2010.143883.
40. Tonne JM, Campbell JM, Cataliotti A, Ohmine S, Thatava T, Sakuma T, et al. Secretion of glycosylated pro-B-type natriuretic peptide from normal cardiomyocytes. *Clin Chem*. 2011;57:864–873. doi: 10.1373/clinchem.2010.157438.
41. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, et al; PROCARDIS Consortium. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med*. 2009;361:2518–2528. doi: 10.1056/NEJMoa0902604.
42. Kanaji T, Okamura T, Osaki K, Kuroiwa M, Shimoda K, Hamasaki N, et al. A common genetic polymorphism (46 C to T substitution) in the 5'-untranslated region of the coagulation factor XII gene is associated with low translation efficiency and decrease in plasma factor XII level. *Blood*. 1998;91:2010–2014.
43. Sheng J, Luo C, Jiang Y, Hinds PW, Xu Z, Hu GF. Transcription of angiogenin and ribonuclease 4 is regulated by RNA polymerase III elements and a CCCTC binding factor (CTCF)-dependent intragenic chromatin loop. *J Biol Chem*. 2014;289:12520–12534. doi: 10.1074/jbc.M114.551762.
44. Bhoola KD, Figueroa CD, Worthy K. Bioregulation of kinins: kallikreins, kininogens, and kininases. *Pharmacol Rev*. 1992;44:1–80.
45. Portelli MA, Siedlinski M, Stewart CE, Postma DS, Nieuwenhuis MA, Vonk JM, et al. Genome-wide protein QTL mapping identifies human plasma kallikrein as a post-translational regulator of serum uPAR levels. *FASEB J*. 2014;28:923–934. doi: 10.1096/fj.13-240879.

CLINICAL PERSPECTIVE

Cardiovascular diseases, including coronary artery disease and venous thromboembolism, are the leading cause of death worldwide. Biomarkers are important tools to diagnose or measure risk of disease, but the causal relationship between biomarkers and diseases is often not clear. Genetic variants that affect levels of protein biomarkers could be used to examine causal relationships between biomarkers and diseases and to provide mechanistic insight into disease. In this study, we investigated whether genetic variants were associated with the levels of 51 serum proteins, 17 of which we had previously identified as predictors for myocardial infarction in the Tromsø Study. We analyzed genotype data from exome sequencing and exome arrays and investigated whether common and rare genetic variation located near the gene (*cis*) that coded for each protein was associated with protein levels. We identified 13 proteins associated with common *cis* variants and 7 proteins associated with rare *cis* variation; 8 of these proteins we had previously identified as biomarkers. To identify pathway-level regulation, we tested whether these significantly associated *cis* variants were also associated in *trans* with the levels of the other 50 proteins in this study. We identified that genetic variation affecting the levels of kallikrein, a protease involved in coagulation, also affect the levels of NTproBNP, a known biomarker for heart failure. We experimentally show that kallikrein can cleave proBNP into NTproBNP and BNP. Our study shows that identifying genetic variants that affect protein levels can provide novel insights and expand our knowledge of the mechanisms of disease.

Circulation
Cardiovascular Genetics

Associations Between Common and Rare Exonic Genetic Variants and Serum Levels of 20 Cardiovascular-Related Proteins: The Tromsø Study

Terry Solomon, Erin N. Smith, Hiroko Matsui, Sigrid K. Braekkan, INVENT Consortium, Tom Wilsgaard, Inger Njølstad, Ellisiv B. Mathiesen, John-Bjarne Hansen and Kelly A. Frazer

Circ Cardiovasc Genet. 2016;9:375-383; originally published online June 21, 2016;
doi: 10.1161/CIRCGENETICS.115.001327
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1942-325X. Online ISSN: 1942-3268

SUPPLEMENTAL MATERIAL

Supplemental Methods

Protein Assay

Protein levels were determined using ELISAs and performed by Tethys Biosciences, Inc (Emeryville, CA), as previously reported[1]. Normal Human Serum from VWR (Radnor, PA), a pool made from 10-16% of Tromsø study samples, and dilution buffer were used as controls. Each anti-protein antibody was either directly conjugated to an AlexaFluor 647 or was biotinylated and detected with a streptavidin-conjugated AlexaFluor 647. Each protein underwent 8 serial dilutions. All samples were performed in triplicate. The eight-point standard curve was measured in six replicates per plate. The AlexaFluor 647-labeled antibodies were detected using the Erenna System (Singulex, Inc., Alameda, CA). Emission from each labeled antibody is measured with a photon detector. The photon detector transmits an electronic pulse for each photon detected, and pulses are counted in 1-ms bins. Binned pulses that exceed a six standard deviation threshold above background are counted. Pulses are recorded as photons/minute. For each protein, it must be detected in >70% of samples, there must be more than 2 logs of standard curve linear range in the ELISA, and there was less than 20% of variance between within-plate replicates for the assay to be considered successful.

Variant Identification and Annotation

Genotypes were determined using either the Illumina Infinium HD HumanExome BeadChip (N=87) or whole-exome sequencing (N=243) using Agilent SureSelect 50 Mb or V4 capture kits and Illumina TruSeq paired-end 100bp cluster kits. Sequence reads were mapped to the reference human genome (hg19) using BWA (version 0.7.10-r789) with default parameters and then

processed using Picard (version 1.115, tool Mark Duplicates) (http://broadinstitute.github.io/picard) and GATK (version 3.3-0, tools RealignerTargetCreator, IndelRealigner, BaseRecalibrator, PrintReads, and HaplotypeCaller). We previously showed that the concordance of the exome sequencing and array genotyping data used in this study is 99.33%[2], therefore we felt confident that we could combine the genotypes from both platforms. Using the array data or information from both on and off-target reads[3] from the sequencing data, genotypes were imputed to the whole genome using Beagle (version 4.0, r1398) and haplotypes from unrelated individuals from the European (EUR) and East Asian (EAS) superpopulations of the 1000 Genomes Project Phase 3[4] for sites with a combined MAF >1%. Due to the difference in coverage and imputation quality between sites that were exome sequenced or assayed by array, we only used sites that had a call rate of >90% in their respective datasets, and then added in additional imputation sites passing QC thresholds (allelic r^2 of >0.3). These two datasets were combined to get a final VCF with imputed and genotyped sites for both exome sequenced and exome genotyped individuals. Because the Tromsø Study is a population-based cohort study, it naturally includes some proportion of related individuals. Of the 330 individuals assayed, 20 were related to another individual in the study at an identity-by-descent value of 0.1 for exome sequenced individuals or 0.2 for arrayed individuals, based on genome-wide data.

All significant common variants were annotated for functional effects using variant effect predictor (VEP)[5], RefSeq genes, the hg19 reference genome, GeneVisible (http://genevisible.com/search) and ROADMAP[6] data of the 28-state chromHMM for Liver (E066), HepG2 (E118), and Monocyte (E029) cells. All rare variants (MAF≤5%) were annotated using VEP, RefSeq, and hg19.

Power Calculations

Our power to detect causal variants of varying effect sizes was determined using an equation from the Abecasis laboratory (http://genome.sph.umich.edu/wiki/Power_Calculations:_Quantitative_Traits) for common variants and the SKAT R package[7] for rare variants. Power for the common variant analyses was calculated using a sample size of 300 individuals, a phenotypic variance (R^2) from 0.0 to 1.0, and alpha levels of 6.97×10^{-7} for the *cis* association, 7.29×10^{-5} for the *cis*-acting-in-*trans* analysis, and 1.25×10^{-8} for the *trans* analysis (Supplemental Table 4, Supplemental Figure 1). Power for the rare variant analysis was calculated using a sample size of 300 individuals, an effect size (β in standard deviations) from 0.0 to 5.0, the default haplotypes (European) for the SKAT package, a causal MAF cutoff of 5%, a sampling subregion length of 3kb, and alpha levels of 3.72×10^{-4} for the *cis* association, 5.30×10^{-5} for the *cis*-acting-in-*trans*, and 9.21×10^{-6} for the *trans* associations (Supplemental Table 4, Supplemental Figure 2).

Supplemental References:

1. Wilsgaard T, Mathiesen EB, Patwardhan A, Rowe MW, Schirmer H, Løchen ML, et al. Clinically significant novel biomarkers for prediction of first ever myocardial infarction: the tromsø study. *Circ Cardiovasc Genet*. 2015;8:363-71.

2. Carson AR, Smith EN, Matsui H, Brækkan SK, Jepsen K, Hansen J, et al. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics*. 2014;15:125.

3. Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, et al. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat Genet*. 2012;44:631-5.

4. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56-65.

5. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P and Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010;26:2069-70.

6. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010;28:1045-8.

7. Wu MC, Lee S, Cai T, Li Y, Boehnke M and Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011;89:82-93.

8. Kim S, Swaminathan S, Inlow M, Risacher SL, Nho K, Shen L, et al. Influence of genetic variation on plasma protein levels in older adults using a multi-analyte panel. *PLoS One*. 2013;8:e70269.

9. Johansson Å, Enroth S, Palmblad M, Deelder AM, Bergquist J and Gyllensten U. Identification of genetic variants influencing the human plasma proteome. *Proc Natl Acad Sci U S A*. 2013;110:4673-8.

10. Lourdusamy A, Newhouse S, Lunnon K, Proitsi P, Powell J, Hodges A, et al. Identification of cis-regulatory variation influencing protein abundance levels in human plasma. *Hum Mol Genet*. 2012;21:3719-26.

11. Liu Y, Buil A, Collins BC, Gillet LC, Blum LC, Cheng LY, et al. Quantitative variability of 342 plasma proteins in a human twin population. *Mol Syst Biol*. 2015;11:786.

12. Kyriakou T, Seedorf U, Goel A, Hopewell JC, Clarke R, Watkins H, et al. A common LPA null allele associates with lower lipoprotein(a) levels and coronary artery disease risk. *Arterioscler Thromb Vasc Biol*. 2014;34:2095-9.

13. Zhu C, Odeberg J, Hamsten A and Eriksson P. Allele-specific MMP-3 transcription under in vivo conditions. *Biochem Biophys Res Commun*. 2006;348:1150-6.

Supplemental Table 1: Cohort statistics

	Sex	N	Age	BMI	N Exome Sequenced	N Exome Arrayed
VTE Controls	Females	64	65.29 (50-74)	26.40 (17.6-38.6)	45 (70%)	19 (30%)
	Males	70	64.14 (46-74)	25.85 (20.3-34.6)	52 (74%)	18 (26%)
VTE Cases	Females	100	63.71 (50-75)	27.21 (18.6-41.8)	74 (74%)	26 (26%)
	Males	96	63.45 (45-74)	26.94 (19.7-36.4)	72 (75%)	24 (25%)
Total		330	64.03 (45-75)	26.69 (17.6-41.8)	243 (74%)	87 (26%)

N, number

Supplemental Table 2: The fifty-one proteins and 50 loci (C3 and C3b derive from the same gene and locus, but are considered two proteins here) used in this study.

Protein Symbol	Protein Name	Gene	Chr	Locus Start	Locus End	N *cis* Variants	N Individuals Measured	Phenotypic Mean	Range (min-max)	Units
a2-AP	alpha-2-antiplasmin	*SERPINF2*	17	1146129	2158559	1295	303	3.06	0.07-17.79	ug/mL
ACE	angiotensin-converting enzyme	*ACE*	17	61054421	62075741	1075	318	858.45	15.51-3344.92	ng/mL
ADIPOQ	adiponectin	*ADIPOQ*	3	186060462	187076252	766	329	6.81	0.74-99.32	ug/mL
AGER	advanced glycosylation end product-specific receptor	*AGER*	6	31648744	32652099	3523	303	0.40	0.08-1.67	ng/mL
AGT	angiotensinogen	*AGT*	1	230338271	231350336	678	330	1.37	0.01-13.71	ug/mL
ANG	angiogenin	*ANG*	14	20652335	21662345	1516	298	219.78	89.63-482.54	ng/mL
APOA1	apolipoprotein A-1	*APOA1*	11	116206468	117208338	721	302	1582.73	0.08-3704.94	ug/mL
APOB	apolipoprotein B-100	*APOB*	2	20724300	21766945	482	303	19.44	3.82-189.05	ug/mL
APOC3	apolipoprotein C-III	*APOC3*	11	116200623	117203787	715	294	255.52	108.25-960.93	ug/mL
BGLAP	osteocalcin	*BGLAP*	1	155711950	156713123	1564	328	6.27	0.60-63.58	ug/mL
BSG	basigin	*BSG*	19	71324	1083493	2306	328	42.99	15.03-87.72	ng/mL
C3	complement C3	*C3*	19	6177845	7220662	1714	284	233.77	71.76-631.00	mg/mL
C3b	complement component C3b	*C3*	19	6177845	7220662	1714	301	2.95	1.05-16.21	ug/mL
CCL5	C-C motif chemokine 5	*CCL5*	17	33698495	34707377	923	328	155.92	7.77-759.97	ng/mL
CD14	monocyte differentiation antigen CD14	*CD14*	5	139511312	140513286	1104	328	244.71	76.28-739.07	ng/mL
CD163	scavenger receptor cysteine-rich type 1 protein M130	*CD163*	12	7123411	8156414	930	328	180.84	14.71-4863.99	ng/mL
CD40-L	CD40 ligand	*CD40LG*	X	135230335	136242549	441	303	13.31	1.84-154.09	ng/mL
CHIT1	chitotriosidase-1	*CHIT1*	1	202685206	203698860	951	303	57.02	0.27-477.70	ng/mL
CRP	C-reactive protein	*CRP*	1	159182078	160184379	1050	303	199.32	2.50-4877.52	ng/mL
CST3	cystatin-C	*CST3*	20	23114293	24118574	459	302	678.73	239.86-2412.46	ng/mL
CTSG	cathepsin G	*CTSG*	14	24542723	25545466	1539	330	37.26	6.88-186.27	ng/mL
CXCL10	C-X-C motif chemokine 10	*CXCL10*	4	76442268	77444689	985	303	0.05	0.01-0.74	ng/mL
DCN	decorin	*DCN*	12	91039034	92073359	128	303	13.47	6.00-31.13	ng/mL
DPP4	dipeptidyl peptidase 4	*DPP4*	2	162348754	163431052	530	328	992.83	225.56-5879.35	ng/mL
F12	coagulation factor XII	*F12*	5	176329138	177336577	955	303	24.82	0.94-50.27	ug/mL
Fetuin A	alpha-2-HS-glycoprotein	*AHSG*	3	185830849	186839107	752	302	924.06	401.88-3961.18	ug/mL

FTH1	ferritin heavy chain	*FTH1*	11	61231756	62235132	918	329	286.80	6.05-4905.20	ng/mL
HP	haptoglobin	*HP*	16	71588507	72594955	1000	301	799.22	0.06-4090.92	ug/mL
HSPA1B	heat shock 70kDa protein 1B	*HSPA1B*	6	31295511	32298031	3372	303	4.05	0.90-62.02	ng/mL
ICAM1	intercellular adhesion molecule 1	*ICAM1*	19	9881516	10897291	1920	328	34.59	4.31-97.50	ng/mL
KLKB1	plasma kallikrein	*KLKB1*	4	186648671	187679625	798	303	29494.68	2.46-59991.93	ng/mL
KNG1	kininogen-1	*KNG1*	3	185935097	186960678	809	302	84574.02	5.57-302245.75	ng/mL
LBP	lipopolysaccharide-binding protein	*LBP*	20	36474884	37505653	953	303	984.41	170.31-2056.94	ng/mL
LP(a)	apolipoprotein(A)	*APOA*	6	160452514	161587407	1080	303	343.47	12.11-3656.08	ng/mL
MMP3	stromelysin-1	*MMP3*	11	102206527	103214342	1056	328	11.26	2.21-75.17	ng/mL
MMP8	neutrophil collagenase	*MMP8*	11	102082525	103095685	971	302	14.34	1.47-68.09	ng/mL
MMP9	matrix metalloproteinase-9	*MMP9*	20	44137546	45145200	1312	303	458.41	106.99-2665.14	ng/mL
MPO	myeloperoxidase	*MPO*	17	55847216	56858296	1116	303	60.57	12.77-271.17	ng/mL
NTproBNP	n-terminus pro-brain natriuretic protein	*NPPB*	1	11417520	12418992	1380	303	0.30	0.02-6.38	ng/mL
PAI-1	plasminogen activator inhibitor 1	*SERPINE1*	7	100270378	101282547	2129	301	38.30	15.35-95.10	ng/mL
REN	renin	*REN*	1	203623943	204635465	1018	329	0.63	0.05-3.33	ng/mL
SHBG	sex hormone-binding globulin	*SHBG*	17	7017381	8036700	2576	303	1.79	0.21-7.28	ug/mL
TAFI	carboxypeptidase B2	*CPB2*	13	46127321	47179211	498	303	27.92	15.36-68.45	ug/mL
THBS1	thrombospondin-1	*THBS1*	15	39373279	40389668	413	320	44.55	4.05-190.52	ug/mL
THBS4	thrombospondin-4	*THBS4*	5	78831169	79879107	810	303	2.79	0.16-192.66	ug/mL
TIMP1	metalloproteinase inhibitor 1	*TIMP1*	X	46941689	47946190	501	328	61.42	7.27-176.27	ng/mL
TIMP4	metalloproteinase inhibitor 4	*TIMP4*	3	11694567	12700851	573	302	5.20	1.88-18.00	ng/mL
TNFRSF11B	tumor necrosis factor receptor superfamily member 11B	*TNFRSF11B*	8	119435795	120464383	175	328	49.80	11.71-152.77	ng/mL
TNFRSF1B	tumor necrosis factor receptor superfamily member 1B	*TNFRSF1B*	1	11727059	12769277	1355	328	19.79	7.56-49.06	ng/mL
uPAR	urokinase plasminogen activator surface receptor	*PLAUR*	19	43650246	44674498	1362	303	1.73	0.54-4.91	ng/mL
VCAM1	vascular cell adhesion protein 1	*VCAM1*	1	100685195	101704601	326	328	138.01	39.81-679.86	ng/mL

Supplemental Table 3: Number of tests performed for each type of association analysis and the P-value cutoffs using Bonferroni correction or permutations for a FWER < 0.05.

Variant Class	Analysis	Number of tests	Bonferroni cutoff	Permutation cutoff
Common, single site	*Cis*	100,378	4.98×10^{-7}	6.91×10^{-7}
	Cis-acting-in-*trans*	663	7.40×10^{-5}	7.29×10^{-5}
	trans	5,119,278	9.77×10^{-9}	1.25×10^{-8}
Rare, collapsed	*Cis*	153	3.21×10^{-4}	3.72×10^{-4}
	Cis-acting-in-*trans*	918	5.34×10^{-5}	5.30×10^{-5}
	trans	7,803	6.16×10^{-6}	9.21×10^{-6}

Supplemental Table 4: Amount of phenotypic variance explained (R^2) and effect size (β) detected for the various analyses when there is 80% power.

Variant Class	Analysis	Alpha	Variance Explained (R^2)
Common, single-site	*cis*	6.91×10^{-7}	0.113
	cis-as-*trans*	7.29×10^{-5}	0.078
	trans	1.25×10^{-8}	0.143
Variant Class	**Analysis**	**Alpha**	**Effect Size (β)**
Rare, collapsed [100% causal]	*cis*	3.72×10^{-4}	0.80
	cis-as-*trans*	5.30×10^{-5}	0.90
	trans	9.21×10^{-6}	1.0
Rare, collapsed [50% causal]	*cis*	3.72×10^{-4}	1.25
	cis-as-*trans*	5.30×10^{-5}	1.45
	trans	9.21×10^{-6}	1.75

Supplemental Table 5: Number of variants that were directly genotyped or imputed for the exome sequenced and exome arrayed individuals.

	Exome Sequenced (N = 243)	**Exome Arrayed (N = 87)**	**Combined**
Genotyped	24,008	2,563	24,915
Imputed	129,502	56,195	138,415
Total	153,510	58,758	158,137

N, number of individuals

Supplemental Table 6: Type of variants that were directly genotyped or imputed for the exome sequenced and exome arrayed individuals.

Variant type	Exome Sequenced (Genotyped)	Exome Arrayed (Genotyped)	Exome Sequenced (Imputed)	Exome Arrayed (Imputed)	% of total Exome Sequenced variants that were imputed	% of total Exome Arrayed variants that were imputed
Intergenic	1161	355	66432	25886	98.3%	98.6%
Non-coding RNA	694	30	783	470	53.0%	94.0%
Intronic	14700	459	59702	27640	80.2%	98.4%
UTR variant	1323	53	2017	1168	60.4%	95.7%
Synonymous	2488	73	203	600	7.5%	89.2%
Missense	3378	1568	313	380	8.5%	19.5%
Coding sequence indels	192	0	16	34	7.7%	100.0%
start or stop related variants	72	25	7	8	8.9%	24.2%

Supplemental Table 7: Reported disease associations of the significant *cis*-pQTLs.

Locus	Top Variant	Chr	Position	Known pQTL? (if reported direction of effect matches this study's direction of effect)	Schadt eQTL	GTEx eQTL	Genome-wide associations	OMIM
AGT	rs4762	1	230845977	Kim[8] (yes)		AGT sun-exposed skin, transformed fibroblasts, non-exposed skin, subcutaneous adipose, lung, colon, breast, testis, esophagus (2.0x10-6 to 1.3x10-33); RP11-99J16_A.2 transformed fibroblasts, sun-exposed skin, subcutaneous adipose (6.0x10-10 to 1.0x10-14)		in LD with rs699 and rs5051 (r^2 = 0.24, D' = 1) which are associated with hypertension, lower promoter activity and transcription amount
ANG	rs3748338	14	21167576	novel	rs8008440 (r^2=0.24) -log10p = 5.7144			
C3	rs11569415	19	6716279	Johansson[9] (yes)				in LD with rs2230199 (r^2 = 0.98, D' = 0.99) which is associated with age-related macular degeneracy
C3	rs2230199	19	6718387	Johansson[9] (yes)			age-related macular degeneration	age-related macular degeneracy, slow and fast in electrophoresis, kidney production of C3
CHIT1	rs2486951	1	203174921	Lourdusamy[10] (yes)		CHIT1 whole blood (4.2x10-8); ADORA1 transformed fibroblasts (1.4x10-6)		
F12	rs1801020	5	176836532	Liu[11] (yes)		F12 liver (2.3x10-10); MXD3 esophagus (4.2x10-6)	F12 levels, protective effect on acute coronary syndrome in people with stable CAD, activated partial thromplastin time, serum metabolite levels	

KLKB1	rs3733402	4	187158034	novel	F11 artery, esophagus, brain, and muscle (1.8-7.0x10-6)	B-type natriuretic peptide, midregional-proadrenomedullin and C-terminal-pro-endothelin-1, serum metabolite levels	PKD Sedi is two mutations (compound heterozygosity with rs121964952) that reduce binding to HMWK
KNG1	rs166479	3	186443250	novel		activated partial thromboplastin time	
LBP	rs2232613	20	36997655	Lourdusamy[10] (yes)		reduced binding capacity for LPS, homozygous has low serum concentrations, protease cleavage site, carriers have cleaved LBP which doesn't bind LPS and has low cytokine after LPS	
APOA	rs41272114	6	161006077	Kyriakou[12] (yes)		plasma plasminogen levels	
APOA	rs56393506	6	161089307	Novel			
MMP3	rs7926920	11	102698724	Zhu[13] (yes)	MMP1 transformed fibroblasts (6.8x10-9); WTAPP1 transformed fibroblasts, testis (1.9-8.0x10-7)	Serum MMP-1 levels	
MMP8	rs35231465	11	102584135	novel			
SERPINF2	rs2070863	17	1648502	novel	SERPINF2 skeletal muscle, sun-exposed skin, subcutaneous adipose, transformed fibroblasts, esophagus, testis (5.3x10-7 to 8.8x10-18); WRD81 esophagus, sun-exposed skin, breast (2x10-6 to 1.3x10-12)		

Supplemental Table 8: Functional annotations of the 14 significant *cis*-pQTLs using GeneVisble, variant effect predictor (VEP) and ROADMAP data of the 28-state chromHMM for Monocyte (E029), Liver (E066) and HepG2 (E118) cells.

Protein	Gene	Top Variant	Chr	Postion	Expressed in	macrophage ROADMAP	Liver ROADMAP	HepG2 ROADMAP	VEP Annotations
AGT	*AGT*	rs4762	1	230845977	liver	quiescent	downstream promoter TSS 2	Transcribed & regulatory	missense variant (*AGT*); TF binding site variant (Nrsf)
ANG	*ANG*	rs3748338	14	21167576	liver	quiescent	Weak transcription	quiescent	missense (*RNASE4*)
LP(a)	*APOA*	rs41272114	6	161006077	liver	quiescent	quiescent	quiescent	splice donor variant (*APOA*)
LP(a)	*APOA*	rs56393506	6	161089307		quiescent	quiescent	quiescent	upstream gene variant (*APOA*); regulatory region variant (CTCF binding site)
C3	*C3*	rs11569415	19	6716279	liver	quiescent	quiescent	Weak transcription	intronic (*C3*)
C3b	*C3*	rs2230199	19	6718387	liver	quiescent	downstream promoter TSS 2	downstream promoter TSS 2	missense (*C3*); TF binding site variant (Egr1)
CHIT1	*CHIT1*	rs2486951	1	203174921	macrophage	quiescent	quiescent	quiescent	regulatory region variant (promoter flanking region); intergenic
F12	*F12*	rs1801020	5	176836532	liver	quiescent	Poised promoter	Active transcription start site	5' UTR variant (*F12*)
KLKB1	*KLKB1*	rs3733402	4	187158034	liver	quiescent	quiescent	quiescent	missense (*KLKB1*)
KNG1	*KNG1*	rs1656921	3	186442833	liver	quiescent	quiescent	quiescent	intronic (*KNG1*)
LBP	*LBP*	rs2232613	20	36997655	liver	quiescent	Weak transcription	quiescent	missense (*LBP*)
MMP3	*MMP3*	rs2155013	11	102701858	joints	quiescent	quiescent	quiescent	intronic (*WTAPP1*)
MMP8	*MMP8*	rs35231465	11	102584135	bone marrow	quiescent	quiescent	quiescent	Stop gain (*MMP8*); 3' UTR variant (*MMP8*)
a2-AP	*SERPINF2*	rs8077638	17	1640793	liver	Transcribed 3' preferential	Transcribed 3' preferential	Transcribed 3' preferential	synonymous variant (*WDR81*); upstream gene variant; downstream gene variant

Supplemental Table 9: List of rare variants that comprise each significant rare *cis*-pQTL association and their P-values from the single-site associations. An X means that the variant was used in the indicated clustering method. Gray regions mean that the collapsed region was not significant using that clustering method.

Protein	Chr	Start	rsID	Ref/Effect	MAF	Single-site P-value	MAF ≤ 5%	Deleterious	CADD10
AGER	6	32147044	rs142802704	AAG/A	0.03477	0.0005711	X		
	6	32147157	rs41268928,rs116420335	G/C	0.03465	0.0005482	X		
	6	32148724	rs41270464,rs114878357	C/T	0.04125	0.8907	X		
	6	32148814	.	GGGTTATACAGGAGAGA/G	0.0022	NA	X		
	6	32148909	rs201575255	CTG/C	0.00165	NA	X		
	6	32149065	rs181811810	C/T	0.00441	NA	X		
	6	32149140	rs3176931	C/T	0.00658	NA	X		
	6	32149471	rs114564020	G/A	0.0022	NA	X		
	6	32149571	.	C/A	0.01322	0.01551	X		
	6	32149801	rs9391855,rs116515025	C/T	0.02211	0.03112	X		
	6	32149883	rs204996,rs116334026	C/T	0.0297	0.1413	X		
	6	32150047	rs77170610	C/T	0.00704	NA	X		
	6	32150107	.	TGAGGCCCTATCTCAGG/T	0.0022	NA	X		
	6	32150303	rs144335694	T/C	0.0022	NA	X		
	6	32150523	.	T/C	0.0022	NA	X		
	6	32150872	.	G/A	0.0022	NA	X		
	6	32151443	rs2070600,rs114177847	C/T	0.0396	0.001871	X		
	6	32151458	rs80096349,rs116828224	G/A	0.00165	NA	X		
	6	32151539	.	G/A	0.0022	NA	X		
	6	32151882	rs115111668	C/T	0.0022	NA	X		
Fetuin A	3	186329282	rs111451953	G/A	0.03084	0.4621	X		
	3	186330883	.	G/A	0.00441	NA	X		
	3	186331119	.	T/A	0.00441	NA	X		
	3	186331138	rs150486317	C/T	0.01159	0.5768	X	X	
	3	186331245	rs190631595	G/T	0.0022	NA	X		
	3	186331298	.	A/C	0.00231	NA	X		
	3	186331299	.	G/C	0.00231	NA	X		
	3	186333378	.	C/A	0.0022	NA	X		
	3	186334343	rs79747711	T/G	0.00441	NA	X		
	3	186334932	.	A/G	0.0022	NA	X		
	3	186335056	rs140827890	G/A	0.00661	0.5564	X	X	X
	3	186335248	rs144616056	G/A	0.01762	0.5311	X		
	3	186337746	rs149819140	T/C	0.0022	NA	X		
	3	186337871	rs184392275	G/A	0.0022	NA	X		
	3	186338320	.	T/G	0.0022	NA	X		
	3	186338540	rs35799453	T/C	0.00441	NA	X		

	3	186338564	rs35457250	C/T	0.01821	1.48E-06	X	X	X
	3	186338869	rs11540663	C/T	0.01542	0.9813	X		
CD40LG	X	135741275	.	G/A	0.0022	NA		X	
	X	135741443	rs148594123	G/A	0.0165	0.001338		X	
CHIT1	1	203183825	rs2015402	G/A	0.04846	0.2873	X		
	1	203184018	rs946849	T/C	0.04846	0.2873	X		
	1	203184924	rs80241012	G/A	0.02643	0.1151	X		
	1	203185118	rs17532442	C/T	0.03744	0.2581	X		
	1	203186420	.	A/C	0.01106	0.667	X		
	1	203186666	rs41308417	C/G	0.02212	0.6183	X		
	1	203188379	.	CCCACTGGTTGTCCCGGAAGATGTAGGGCA/C	0.00661	0.02734	X		
	1	203189093	rs76499133	C/T	0.03111	0.02956	X		
	1	203189350	rs74969659	G/A	0.03304	0.2992	X		
	1	203189634	rs2486959	A/G	0.04626	3.46E-13	X		
	1	203191527	rs56152830	C/T	0.01322	0.4565	X		
	1	203191994	rs7512820	G/A	0.00889	0.3586	X		
	1	203192424	rs181385947	G/A	0.02643	0.9288	X		
	1	203192518	.	G/A	0.00661	0.9764	X		
	1	203193134	rs2486068	T/G	0.04405	1.32E-09	X		
	1	203194544	rs140940634	C/CATT	0.03965	0.5655	X		
	1	203194548	rs184545416	G/T	0.03965	0.5655	X		
	1	203194688	.	C/T	0.00441	NA	X		
	1	203194834	rs137852607	C/T	0.0033	NA	X		
	1	203195006	rs116389839	G/A	0.01322	0.2128	X		
	1	203195126	rs185483258	C/T	0.00229	NA	X		
	1	203195398	rs10920587	C/A	0.00881	0.9725	X		
	1	203195689	rs2486070	G/A	0.04405	9.21E-11	X		
	1	203196479	rs2486071	A/C	0.04626	9.39E-11	X		
	1	203196842	rs56035601	C/T	0.03965	0.5655	X		
	1	203198596	rs72739588	G/A	0.02477	0.2938	X		
LP(a)	6	160952621	rs41266381	A/C	0.00441	NA	X		
	6	160952667	rs73012273	C/G	0.01982	0.7021	X		
	6	160952780	rs186413938	C/T	0.0033	NA	X	X	X
	6	160952816	rs41267807	T/C	0.0132	0.4979	X	X	X
	6	160953642	rs41267809	A/G	0.0198	0.01013	X	X	
	6	160960892	.	AAG/A	0.0022	NA	X		
	6	160961137	rs3798220	T/C	0.00828	0.1252	X	X	X
	6	160962115	.	A/G	0.0022	NA	X		
	6	160962151	.	T/G	0.0022	NA	X	X	X
	6	160962185	.	A/G	0.00441	NA	X		
	6	160962190	.	C/A	0.0022	NA	X	X	X
	6	160962366	.	TC/T	0.02535	0.8887	X		
	6	160962368	rs116039216	G/T	0.02546	0.8848	X		
	6	160962370	rs116089584	G/T	0.02804	0.9047	X		

6	160963576	rs41265940	T/A	0.00221	NA	X		
6	160963648	rs6920765	G/A	0.0022	NA	X		
6	160963964	rs41265934	C/G	0.00441	NA	X		
6	160964135	rs41265930	T/C	0.00661	0.868	X		
6	160966559	rs139145675	G/A	0.00165	NA	X	X	X
6	160968863	rs149574804	C/T	0.01982	0.4082	X		
6	160968968	rs41264848	G/A	0.0022	NA	X		
6	160969075	rs41264844	C/T	0.03084	0.5789	X		
6	160969096	rs4708871	C/T	0.02423	0.2038	X		
6	160969113	rs145989243	G/A	0.0022	NA	X		
6	160971286	rs62441900	C/G	0.03965	0.2143	X		
6	160973905	rs62441901	C/G	0.03965	0.2143	X		
6	160976914	.	CA/C	0.00221	NA	X		
6	160978270	rs184372256	A/G	0.00442	NA	X		
6	160978686	rs149526393	A/G	0.00441	NA	X		
6	160985107	rs145783310	A/G	0.0495	0.7805	X		
6	160985438	rs79563112	C/T	0.0495	0.7805	X		
6	160985526	rs118039278	G/A	0.02632	0.0406	X		
6	160997118	rs74617384	A/T	0.02632	0.0406	X		
6	160998052	rs76602267	G/A	0.00441	NA	X		
6	160998143	.	C/T	0.0022	NA	X		
6	161005610	rs55730499	C/T	0.02632	0.0406	X		
6	161005898	.	C/T	0.00221	NA	X		
6	161005908	rs41272116	ACTT/A	0.00221	NA	X		
6	161006077	rs41272114*	C/T	0.0297	1.05E-08	X	X	X
6	161006084	rs76144756	G/A	0.00441	NA	X	X	X
6	161006105	rs41272112	C/T	0.00165	NA	X	X	
6	161007647	.	G/A	0.0022	NA	X		
6	161010118	rs10455872	A/G	0.01815	0.005497	X		
6	161010546	rs41267815	G/C	0.00441	NA	X		
6	161011907	rs74334585	C/T	0.01101	0.1663	X		
6	161012262	rs144958108	A/G	0.00667	0.8523	X		
6	161015301	rs41271036	A/G	0.01101	4.23E-06	X		
6	161016414	.	A/G	0.0022	NA	X		
6	161020526	rs41270998	A/G	0.0022	NA	X		
6	161020532	.	G/A	0.00661	0.417	X		X
6	161021800	.	G/T	0.00221	NA	X		
6	161022107	rs41259144	C/T	0.00662	0.003408	X	X	X
6	161022108	rs186072375	G/T	0.0022	NA	X		X
6	161025782	.	G/T	0.00455	NA	X		
6	161026197	rs117174672	G/A	0.03084	0.2908	X		
6	161026250	.	G/A	0.0022	NA	X		
6	161027256	.	A/G	0.00231	NA	X		
6	161027287	.	G/C	0.00221	NA	X		
6	161027430	rs144587038	G/T	0.01542	0.1217	X		

	6	161027821	rs75055004	A/T	0.0022	NA	X		
	6	161027895	.	TA/T	0.00457	NA	X		
	6	161032267	rs117949336	C/T	0.00461	NA	X		
	6	161032369	rs151298886	C/T	0.00221	NA	X		
	6	161032401	rs78893353	G/A	0.01106	0.1991	X		
	6	161032413	.	G/C	0.00221	NA	X		
	6	161032497	rs112092923	A/G	0.02212	0.09362	X		
	6	161055824	.	C/A	0.00243	NA	X		
	6	161055876	.	C/T	0.00444	NA	X		
	6	161055880	.	G/A	0.00222	NA	X		
	6	161055942	.	G/C	0.00441	NA	X		
	6	161055968	.	C/A	0.0022	NA	X		
	6	161055991	.	C/T	0.0022	NA	X		
	6	161056129	.	T/TAGA	0.0022	NA	X		
	6	161071476	rs200491482	T/G	0.00165	NA	X	X	X
	6	161071615	.	G/A	0.0022	NA	X		
	6	161071638	.	T/C	0.0022	NA	X		
	6	161087336	rs117643720	T/C	0.0022	NA	X		
	6	161087368	.	G/A	0.00881	0.06095	X		
	6	161087372	rs181060240	T/C	0.01322	0.03437	X		
MMP8	11	102584135	rs35231465*	G/A	0.03642	6.86E-07			X
	11	102585130	.	A/G	0.00224	NA			X
	11	102586142	rs61753779	A/G	0.00166	NA			X
	11	102592160	rs11602288	G/A	0.00661	0.7785			X
	11	102593266	rs112188995	C/T	0.00993	0.1649			X
TAFI	13	46627762	rs145067962	A/T	0.0066	0.003747		X	
	13	46648069	rs140446990	T/C	0.00495	0.0407		X	
	13	46656619	.	C/T	0.0022	NA		X	
TIMP4	3	12195137	.	G/C	0.00221	NA		X	X
	3	12195660	rs140022692	G/A	0.01106	0.1126			X
	3	12198889	.	A/G	0.00221	NA		X	X
	3	12200201	.	C/T	0.00221	NA		X	X
	3	12200210	.	G/A	0.00442	NA		X	X
	3	12200219	.	A/G	0.00221	NA		X	X
	3	12200269	.	C/T	0.00221	NA			X

* = variant was tested in both the common analysis and in the rare analysis.

Supplemental Table 10: Pearson's correlation between the proteins that were identified in the trans-pQTL analysis.

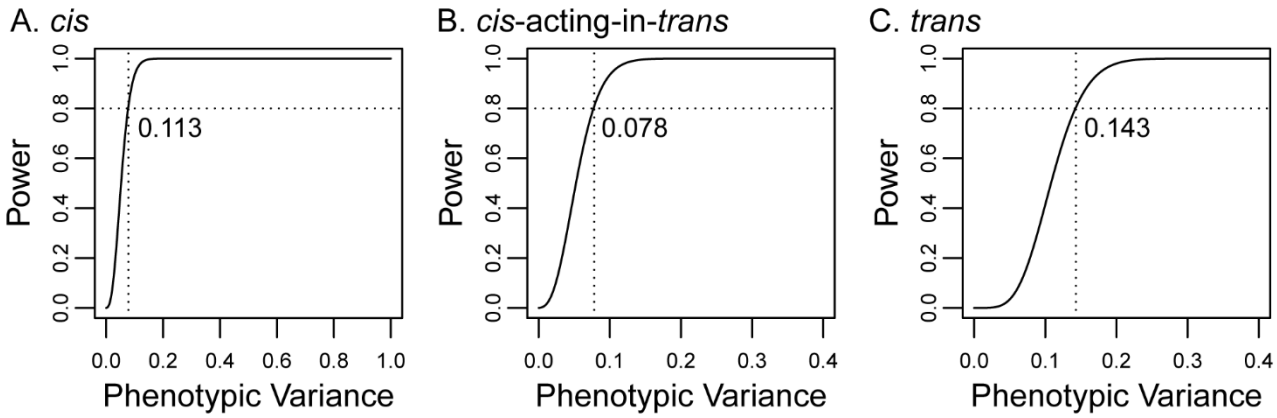
	F12	**KLKB1**	**KNG1**	**NTproBNP**	**uPAR**
F12	1.0	-0.18	0.09	-0.11	-0.06
KLKB1	-0.18	1.0	0.29	0.16	0.20
KNG1	0.09	0.29	1.0	0.09	0.32
NTproBNP	-0.11	0.16	0.09	1.0	0.28
uPAR	-0.06	0.20	0.32	0.28	1.0

Supplemental Table 11: Lookup of common *cis*-pQTLs for their associations in the CARDIoGRAM and INVENT meta-analyses.

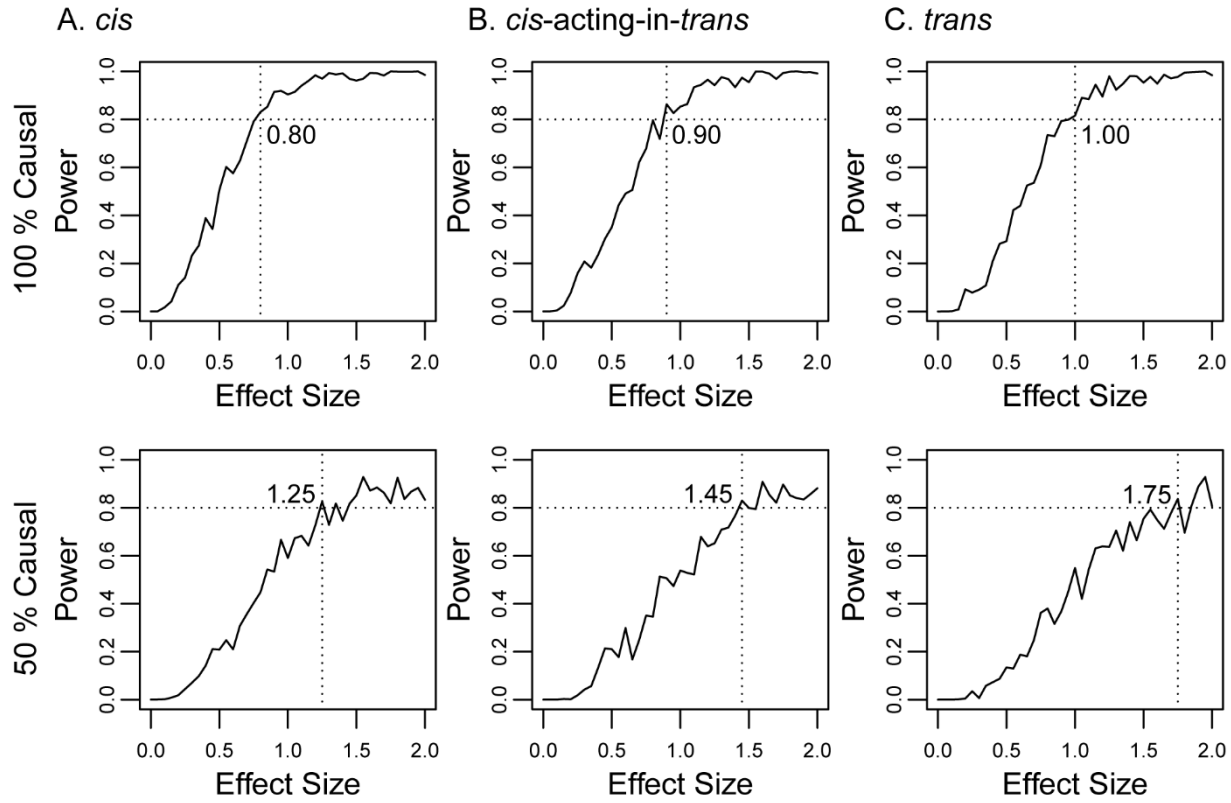
Protein	Variant	Chr	Start	Ref/Alt	CARDIoGRAM Coronary Artery Disease			INVENT Venous Thromboembolism		
					P-value	β	SE	P-value	β	SE
a2-AP	rs8077638*	17	1640793	C/T	0.3399	0.156	0.017	Not in INVENT		
a2-AP	rs8065251*	17	1637458	G/A	0.3552	0.015	0.017	0.3298	-0.028	0.029
AGT	rs4762	1	230845977	G/A	0.5018	-0.143	0.021	0.1694	-0.048	0.035
ANG	rs3748338	14	21167576	A/T	0.4549	0.020	0.027	0.9519	0.002	0.039
C3/C3b	rs2230199†	19	6718387	G/C	0.1045	0.050	0.031	0.7668	0.010	0.034
CHIT1	rs2486951	1	203174921	A/G	0.7288	-0.006	0.017	0.8562	-0.005	0.028
F12	rs1801020	5	176836532	A/G	0.5115	-0.016	0.246	0.6957	0.011	0.028
KLKB1‡	**rs3733402**	4	187158034	G/A	**0.0086**	0.040	0.015	**$8.2x10^{-12}$**	-0.159	0.023
KNG1	**rs166479**	3	186443250	T/C	0.1692	0.020	0.014	**0.0436**	-0.046	0.023
LBP	rs2232613	20	36997655	C/T	0.3624	0.025	0.028	0.3531	0.050	0.054
MMP3	rs7926920	11	102698724	G/A	0.1016	0.023	0.014	0.7347	0.008	0.023

Bolded indicates a nominal P-value of <0.05. β, effect size. SE, standard error
* rs8077638 was not present in the INVENT dataset, so the next most significant variant (rs8065251) was looked up in both studies as well.
† The top variant for C3 (rs11569415) was not present in either CARDIoGRAM or INVENT so the next most significant variant (rs2230199) was used.
‡ The association with VTE was no longer significant after adjustment that included rs4253417.
No significant variants for LP(a) or MMP8 were present in either study.

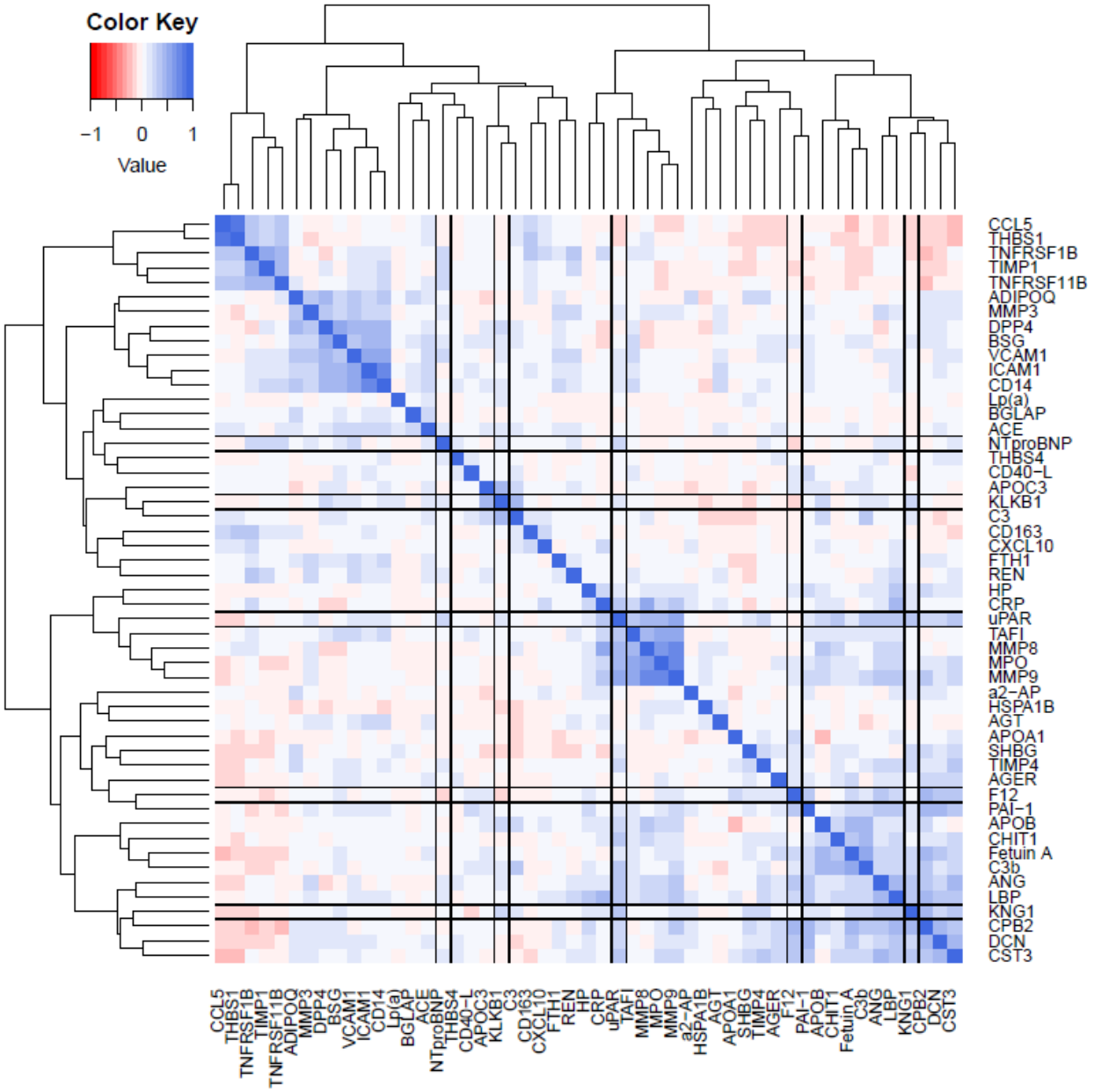
Supplemental Figure 1: Power to detect common variation pQTLs with varying effect sizes in the three stages of analysis. A) Power curve for *cis*-pQTLs using the permutation cutoff of $6.91x10^{-7}$ as the alpha. B) Power curve for testing the *cis*-pQTLs acting-in-*trans* using the permutation cutoff of $7.29x10^{-5}$ as the alpha. C) Power curve for *trans*-pQTLs using the permutation cutoff of $1.25x10^{-8}$. The x-axis is measuring the amount of variance of the phenotype that a variant explains (R^2).



Supplemental Figure 2: Power to detect rare variation pQTLs with varying effect sizes in the three stages of analysis. Effect size is measured in standard deviations (β). The top row assumes that all variants have an equal effect and that all variants are causal. The bottom row assumes that all variants have an equal effect and that half of the variants tested are causal. A) power to detect *cis* associations, alpha = 3.72×10^{-4}; B) power to detect *cis*-acting-in-*trans* pQTLs, alpha = 5.30×10^{-5}; C) power to detect *trans* associations, alpha = 9.21×10^{-6}. The x-axis is measuring the effect size (β) in standard deviations.



Supplemental Figure 3: Pearson's correlation of the protein levels. Dendrogram shows clustering based on correlation. Black lines are outlining the proteins identified in the *cis*-acting-in-*trans* analysis: F12, KLKB1, KNG1, NTproBNP, uPAR; corresponding values can be found in Supplemental Table 10.



Supplemental Figure 4: Silver stain of proBNP incubated for 1 hour with varying concentrations of kallikrein, with and without a kallikrein-specific inhibitor, PPACK II. Kallikrein concentrations are 74.8nM, 374nM, 748nM, and 1497nM. The upper bands are the light and heavy chains of kallikrein. The lower band is proBNP. 374nM of kallikrein was chosen to perform the silver stain and western blot in Figure 4 of the paper.