


Hemostatic Genes Exhibit a High Degree of Allele-Specific Regulation in Liver

Martina Olsson Lindvall¹  Lena Hansson^{2,3} Sofia Klasson¹ Marcela Davila Lopez⁴ Christina Jern¹
Tara M. Stanne¹

¹Department of Laboratory Medicine, Institute of Biomedicine, The Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden

²Novo Nordisk, Oxford, United Kingdom

³Science For Life Laboratory (SciLifeLab), Stockholm, Sweden

⁴Bioinformatics Core Facility, University of Gothenburg, Gothenburg, Sweden

Address for correspondence Tara M. Stanne, PhD, Department of Laboratory Medicine, Institute of Biomedicine, The Sahlgrenska Academy at the University of Gothenburg, Box 445, SE-405 30 Gothenburg, Sweden (e-mail: tara.stanne@gu.se).

Thromb Haemost 2019;119:1072–1083.

Abstract

Objective Elucidating the genetic basis underlying hepatic hemostatic gene expression variability may contribute to unraveling genetic factors contributing to thrombotic or bleeding disorders. We aimed to identify novel *cis*-regulatory variants involved in regulating hemostatic genes by analyzing allele-specific expression (ASE) in human liver samples.


Study Design Biopsies of human liver tissue and blood were collected from adults undergoing liver surgery at the Sahlgrenska University Hospital ($n = 20$). Genomic deoxyribonucleic acid (gDNA) and total ribonucleic acid (RNA) were isolated. A targeted approach was used to enrich and sequence 35 hemostatic genes for single nucleotide polymorphism (SNP) analysis (gDNAseq) and construct individualized genomes for transcript alignment. The allelic ratio of transcripts from targeted RNAseq was determined via ASE analysis. Public expression quantitative trait loci (eQTL) and genome-wide association study (GWAS) data were used to assess novelty and importance of the ASE SNPs (and proxies, $r^2 \geq 0.8$) for relevant traits/diseases.

Results Sixty percent of the genes studied showed allelic imbalance across 53 SNPs. Of these, 7 SNPs were previously validated in liver eQTL studies. For 32 with eQTLs in other cell/tissue types, this is the first time genotype-specific expression is demonstrated in liver, and for 14 ASE SNPs, this is the first ever reported genotype-expression association. A total of 29 ASE SNPs were previously associated with the respective plasma protein levels and 17 ASE SNPs to other relevant GWAS traits including venous thromboembolism, coronary artery disease, and stroke.

Conclusion Our study provides a comprehensive ASE analysis of hemostatic genes and insights into the regulation of hemostatic genes in human liver.

Keywords

- ▶ coagulation factors
- ▶ gene regulation
- ▶ hemostasis
- ▶ single nucleotide polymorphisms
- ▶ thrombosis

 Martina Olsson's ORCID is <https://orcid.org/0000-0002-8609-5607>.

received
January 17, 2019
accepted after revision
March 11, 2019

DOI <https://doi.org/10.1055/s-0039-1687879>.
ISSN 0340-6245.

© 2019 Georg Thieme Verlag KG
Stuttgart · New York

License terms



Introduction

Genetic variation in several genes encoding hemostatic factors has been demonstrated in candidate gene and genome-wide association studies (GWAS) to be associated with circulating levels of these proteins (e.g., fibrinogen¹; von Willebrand factor^{2,3}; tissue-type plasminogen activator⁴; coagulation factor VII [FVII]²; and thrombin activatable fibrinolysis inhibitor [TAFI]⁵) and with the risk of thrombotic events such as ischemic stroke, myocardial infarction, and venous thrombosis (e.g., see Refs. 6–8) and bleeding disorders (e.g., see Refs. 9 and 10). However, the functional deoxyribonucleic acid (DNA) elements through which these variants exert their effects on these traits remain largely unknown. The liver plays a major role in producing proteins that are secreted into the blood, including many factors involved in hemostasis. Altered levels of circulating hemostatic proteins can lead to thrombus formation and/or bleeding. It is therefore important to understand the molecular mechanisms governing the expression of these genes in liver. Expression quantitative trait loci (eQTL) analysis, which links variations in genotypes to gene expression across individuals, is one approach to elucidate whether genetic variants correlate to gene expression. Several databases have been developed to collect eQTLs, including the Genotype-Tissue Expression (GTEx) portal.¹¹ However, these publically available data sets currently have low coverage of eQTLs in hemostatic genes in human liver tissue, for example, GTEx currently has eQTLs reported for just 10 hemostatic genes. Furthermore, results from eQTL analyses are influenced by unmeasured confounding variables, and thus the reproducibility of eQTL studies is relatively low.^{12,13}

Analysis of allele-specific expression (ASE) is an alternative and complementary approach to eQTL analyses to identify *cis*-regulatory variation.¹⁴ ASE occurs when transcription from one allele is selectively silenced or enhanced, or when transcripts undergo selective posttranscriptional degradation (e.g., nonsense-mediated decay). By comparing expression of alleles within each individual, each allele acts as an internal control for confounding factors (e.g., *trans*-acting effects and environmental confounders) that alter the overall expression of that gene.¹⁵ Thus, the ASE approach is a powerful methodology to map putative *cis*-acting polymorphisms even when using a low number of individuals.¹⁵ If allelic imbalance is observed between the two alleles, *cis*-regulatory variation cosegregating with the marker (or the marker itself) is likely to exist.

Owing to the wide dynamic range of the transcriptome, in which a minority of highly expressed genes constitutes the majority of ribonucleic acid (RNA) molecules, whole-transcriptome RNAseq often results in sparse coverage of many transcripts. Targeted RNAseq (also called RNA CaptureSeq) can overcome these challenges by providing greater depth and power to detect ASE compared with whole-transcriptome RNAseq.¹⁶ eQTL and early ASE studies suggest widespread tissue-specific effects of *cis*-regulatory variants,¹⁷ and analyzing the appropriate tissue is therefore paramount. As far as we are aware, there are no studies that have looked at

ASE in hemostatic genes in human liver tissue samples. Given this background, the main aim of the present study is to identify novel putative *cis*-acting variants involved in regulating genes that are important for hemostasis and expressed in the liver by analyzing ASE in human liver using targeted genomic DNA (gDNA) and RNAseq. A further aim is to validate the identified ASE single nucleotide polymorphisms (SNPs) in previous eQTL studies and investigate if they are associated with plasma levels of the respective proteins and/or other relevant disease and nondisease phenotypes in published GWAS data. The 35 genes selected for analysis represent vitamin K-dependent coagulation factors, other coagulation factors, inhibitors of coagulation, contact activation pathway (intrinsic), fibrinolysis, inhibitors of fibrinolysis, and relevant vitamins and cofactors (see **Table 1**).

Methods

Sample Collection

Surgical resections of apparently healthy (macroscopically noncirrhotic and nontumorous) tissue were collected from 20 unrelated adults of European descent undergoing liver surgery at the Sahlgrenska University Hospital. Immediately following resection, the tissue was placed in RNALater at 4°C for 3 to 4 days, then aliquoted and stored at –80°C. Peripheral blood samples from each individual were also collected. Blood samples were anticoagulated with ethylenediaminetetraacetic acid and aliquots were stored at –80°C within 24 hours. Clinical characteristics of study participants are summarized in **Supplementary Table S1** (available in the online version). The institutional review board, that is, the Ethics Committee at the University of Gothenburg, approved

Table 1 The 35 genes included in the custom gDNA- and RNA-sequencing designs

Gene category	Number	Gene name
Vitamin K-dependent coagulation factors	7	F2, F7, F9, F10, PROC, PROZ, PROS1
Other coagulation factors	10	CPB2, F5, F11, F12, F13B, FGA, FGB, FGG, HABP2, SERPINA10
Inhibitors of coagulation	5	A2M, SERPINA1, SERPINC1, SERPING1, TFPI
Contact activation pathway (intrinsic)	2	KLKB1, KNG1
Fibrinolysis	1	PLG
Inhibitors of fibrinolysis	3	SERPINA5, SERPINE1, SERPINF2
Vitamins, cofactors	5	GGCX, HRG, MCFD2, SERPIND1, VKORC1
Other hemostatic genes ^a	2	C4BPB, LMAN1

Abbreviations: gDNA, genomic deoxyribonucleic acid; RNA, ribonucleic acid.

^aAccording to the Molecular Signatures Database (MSigDB, Broad Institute).

the collection of the human liver tissues and blood samples, and their subsequent use for the purposes of this study. Written informed consent was obtained from all patients prior to participation in this study.

Of the 20 samples collected, high-quality DNA was isolated for 20 samples, and 19 of these were successfully genotyped (see the “Methods” section). Of the 19 successfully genotyped samples, high-quality RNA was isolated and successfully profiled on 17 samples. This set of 17 genotyped and expression profiled samples comprised the ASE data set and all results presented are based on these 17 samples. All samples and patient data were handled in accordance with the policies and procedures of the University of Gothenburg.

Targeted Gene Panel

For the custom gene panel design, we selected 35 genes that are expressed in the liver and are important for hemostasis by studying the literature, the Kyoto Encyclopedia of Genes and Genomes pathway database, and the Molecular Signatures Database (Broad Institute; ▶Table 1). Custom capture kits (Agilent, Santa Clara, California, United States) were designed for targeted sequencing (seq) for both gDNA- and RNAseq using Agilent’s SureDesign tool and eArray, respectively. To exploit the full potential of the design, we also included 42 other genes of interest for our group. For gDNA, probes were designed to capture all exons, introns, and all potential upstream ($\geq 5,000$ base pair [bp]) and downstream (500 bp) regulatory regions, resulting in a 3.3M bp design. For messenger RNA (mRNA), exons from all known isoforms were included resulting in a 0.5M bp design.

DNA Extraction and Variant Analysis (DNAseq)

Total gDNA was extracted from whole blood for each participant using the QIAmp DNA Blood Midi Kit (Qiagen, Hilden, Germany). Strand-specific sequencing libraries for gDNAseq were prepared using 3 μ g DNA and the custom-designed SureSelect kits (Agilent) according to the manufacturer’s protocols. Six to seven purified libraries were multiplexed per sequence lane using 150 bp paired-end reads on a MiSeq sequencer (Illumina, San Diego, California, United States), following the manufacturer’s protocols. Approximately 5M reads were generated per sample (for coverage per gene see ▶Supplementary Fig. S1, available in the online version). All raw sequence reads were quality controlled using FastQC (Babraham Bioinformatics, Cambridge, United Kingdom). Reads were trimmed (removal of adaptors and indexes) with Prinseq-lite and aligned to the reference human genome (GRCh37/hg19) with Bowtie2. Genetic variants were called using GATK HaplotypeCaller v3.4.36 with the GATK Best Practice recommendations, using a minimum required quality score of ≥ 40 .

RNA Extraction and Complementary DNA Sequencing (RNAseq)

Total RNA was extracted from liver tissues in biological duplicates (i.e., two separate pieces of liver tissue per individual) using the AllPrep DNA/RNA mini kit (Qiagen). Strand-

specific RNAseq libraries were prepared for each duplicate separately using a custom-designed SureSelect kit (Agilent) according to the manufacturer’s protocol, and sequenced once per duplicate. A total of 10 to 12 sample libraries were multiplexed per sequence lane using 125 bp paired-end reads on a HiSeq sequencer (Illumina). Approximately 25M reads were generated per sample (for coverage per gene see ▶Supplementary Fig. S1, available in the online version). Read counts and normalized expression values were highly replicable between the duplicates ($R > 0.98$ for all 17 individual samples/duplicates). As a positive control, we ensured that all transcripts of interest were indeed identified. A negative control, *HYAL4*, was not detected. Additionally, samples from two patients were also sequenced using Illumina’s TruSeq kit for whole transcriptome analysis to identify possible weaknesses or biases in our custom-designed kit (prepared as above). Assessment of the data showed no bias in expression values between the two kits, but as expected, coverage was much higher in targeted-seq samples (data not shown). Sequence reads were trimmed (removal of adaptors and indexes) using TrimGalore! and aligned via Star Aligner v2.4.2 with default parameters to N-masked individual-specific reference human genome (GRCh37/hg19) using variant information from the gDNAseq. This was to minimize downstream mapping biases for the ASE analyses, as recommended by GATK ASE best practices.¹⁸ Reads were counted using HTSeq v0.6.1 using default parameters.

Detection of ASE

Read counts from biological duplicates were merged prior to ASE analysis. ASE was then determined based on the allelic ratio of heterozygous SNPs in processed RNAseq data (see above). The GATK-HaplotypeCaller v3.4.36 was used to determine zygosity of the processed RNAseq data using a minimum required quality score of ≥ 60 . A binomial test on the proportion of reference alleles versus total counts was performed using the ASEReadCounter v3.4.46 (GATK, Broad Inst, Boston, Massachusetts, United States), following GATK best practices¹⁸ where applicable. For the ASEReadCounter, a read depth of ≥ 10 and a quality score of ≥ 10 were required. Allelic ratios were standardized such that the minor allele is the alternative allele. Variants were classified as ASE sites if at least one individual had an imbalanced allelic ratio of < 0.35 (minor allele preferred) or > 0.65 (major allele preferred) and a mean allelic ratio averaged among all heterozygous individuals of ≤ 0.4 or ≥ 0.6 .

Annotation and Functional Prediction of Variants

Database of Single Nucleotide Polymorphisms (build 150, available from: <http://www.ncbi.nlm.nih.gov/SNP/>) was used to annotate SNPs. HaploReg v4.1¹⁹ was used to determine the chromatin state in liver tissue per haplotype block. The default 15-state core ChromHMM model and H3K4me1/H3K4me3 peaks from Roadmap Epigenomics²⁰ were used to identify enhancers and promoters. For prediction on whether ASE SNPs abolish, create, or change the affinity of one or several transcription factor (TF) binding sites, HaploReg v4.1 and SNP2TFBS²¹ were used.

Overlap of ASE SNPs in Publically Available eQTL

Databases

We queried all ASE variants and their high linkage disequilibrium (LD) proxies ($r^2 \geq 0.8$ in 1000G European panel) against publicly available eQTL data using PhenoScanner (<http://www.phenoscaner.medschl.cam.ac.uk>; assessed February 2019²²) which includes results from the GTEx project v6 (<http://www.gtexportal.org/home/datasets>^{11,23}), NHLBI GRASP v2,²⁴ STARNET,²⁵ and other smaller studies (see <http://www.phenoscaner.medschl.cam.ac.uk/studies/> for details). eQTL results were filtered to retain only variants with $p < 1 \times 10^{-5}$ and manually curated to include only eQTLs correlating to the 35 genes of interest or their antisense long noncoding RNA (lncRNA).

Assessment of ASE SNPs in GWAS

To identify ASE SNPs associated with circulating levels of the respective proteins and relevant disease and nondisease phenotypes, we queried our variants and corresponding proxies ($r^2 \geq 0.8$) against publicly available GWAS data using both the PhenoScanner “Proteins” and “Diseases & traits” search functions (assessed February 2019²²) which includes results from NHGRI-EBI GWAS catalog (<https://www.ebi.ac.uk/gwas/>) and the UK Biobank (<http://www.ukbiobank.ac.uk>). We supplemented this search with the most recent submissions to the NHGRI-EBI GWAS catalog (downloaded February 2019; version 2019–01–11). For GWAS, summary statistics were filtered to $p < 5 \times 10^{-5}$ and then manually curated to retain only entries with relevant traits.

Assessment of Circulating PAI-1 Levels according to rs6092 Genotype

Controls from the Sahlgrenska Academy Study on Ischemic Stroke (SAHLSIS²⁶) were used to investigate whether rs6092 in *SERPINE1* associates to circulating plasma levels of plasminogen activator inhibitor-1 (PAI-1). This SNP was selected for analysis as it has not been reported to associate to PAI-1 protein levels in published GWAS data and both genotype and protein data were available in SAHLSIS. SAHLSIS includes consecutive patients with ischemic stroke and 600 population-based controls. Only control subjects without a history of cardiovascular disease or any other thrombotic disease or signs of cardiovascular disease on electrocardiogram were included (as described in detail in Jood et al²⁶). Measurement of PAI-1 plasma concentration (expressed as ng/mL) and genotyping in SAHLSIS were previously reported.^{27,28} PAI-1 level comparisons between individuals homozygous for the reference and alternative alleles of rs6092 were performed using Mann–Whitney *U* test in SPSS for Windows version 20 (IBM Corporation, New York, United States). The statistical significance level was 0.05 and *p*-values were two-tailed.

Validation and Independent Replication of Selected ASE SNPs

For replication of selected ASE SNPs, liver tissue and blood samples were collected from an additional 48 individuals. DNA and RNA were isolated as described above. Allele-specific genotyping of two ASE SNPs (rs7337140 in *CPB2*

and rs2227424 in *FGB*) and one negative control (rs12937244 in *SERPINA1*) was performed using rhAMP quantitative polymerase chain reaction (qPCR) assays (Integrated DNA Technologies, Coralville, Iowa, United States), following manufacturer’s recommendations. This method uses two allele-specific forward primers to determine ASE through competitive binding. For ASE analysis, normalized mRNA was converted to complementary DNA (cDNA) using TATAA Grandscript cDNA Synthesis kit #A103 (TATAA Biocenter AB, Gothenburg, Sweden) and then genotyped using the same rhAMP assays outlined above for reverse transcription (RT)-qPCR. Allelic ratios of the two alleles were estimated using threshold cycle (C_t) numbers.

Results

Hemostatic Genes Exhibit Widespread Allele-Specific Expression

Twenty-one hemostatic genes (out of 35) showed allelic imbalance in liver across a total of 53 SNPs (►Fig. 1 and ►Tables 2 and 3). A total of 29 SNPs favored the major allele, 23 favored the minor allele, and 1 SNP showed bidirectional allelic imbalance (rs6107: $n = 1$ heterozygous individual favored the minor allele and $n = 6$ the major allele; ►Tables 2 and 3).

Thirty-Nine ASE SNPs were Validated in eQTL Databases; 14 were Novel

First, we queried *cis*-eQTL variants in publicly available liver tissue eQTL association data using PhenoScanner²² to identify SNPs previously associated with gene expression in liver. Four ASE SNPs were identified as eQTLs in liver tissue for their respective gene or antisense lncRNA. These were in genes encoding coagulation FXII (rs1801020), coagulation FVII (rs491098), complement component 4 binding protein (C4BPB, rs3813948), and TAFI (rs17843995). Notably, the same alleles associated with higher expression in liver tissue eQTL studies were expressed at higher levels in our ASE study. Three additional ASE variants were in high LD with liver eQTL variants for their respective genes. These were in *FGB* encoding fibrinogen (rs2227395 and rs2227401) and *CPB2* encoding TAFI (rs3742264). Thus, we validated seven ASE SNPs as liver eQTLs (referred to as “previously validated” in ►Table 2).

While many *cis*-acting genetic variants are likely to be specific for a given tissue, others are common to several tissues.²³ Further, sample size greatly affects eQTL mapping and more readily available tissues tend to have more eQTLs.²³ Thus, we next queried our ASE SNPs (and corresponding proxies) for eQTL associations in all remaining tissues using PhenoScanner. The six validated associations were also identified as eQTLs in other tissues. A further 32 ASE variants (or proxies) were eQTLs for their respective genes in other tissues. These 32 SNPs were therefore previously associated with gene expression, but this is the first time they are shown to associate with expression in the liver (referred to as “novel to liver” in ►Table 2). For a complete list of all ASE SNPs (or proxies) with eQTLs in any tissue, see ►Supplementary Table S2 (available in the online version).

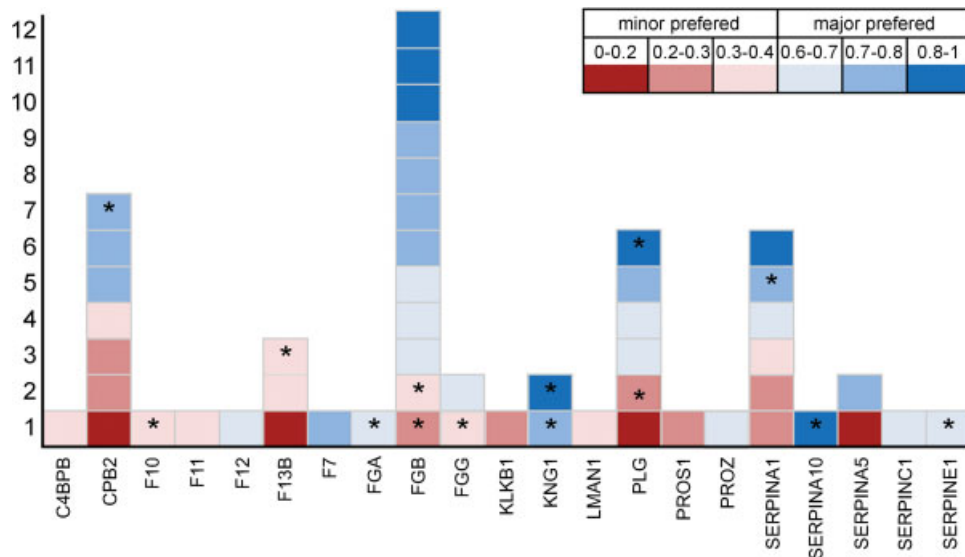


Fig. 1 Twenty-one genes exhibited allele-specific expression (ASE) in human liver. Each gene is shown as vertical bars, with each colored box representing one ASE single nucleotide polymorphism (SNP). The color represents the degree of allelic imbalance averaged among all heterozygous individuals. Red boxes are imbalanced toward the minor allele and blue boxes are imbalanced toward the reference major allele. Darker colors indicate greater level of imbalance. Genes are ordered alphabetically. The majority of SNPs have been externally validated in previously published expression quantitative trait loci (eQTL) data sets ($n = 39$); those novel to our study ($n = 14$) are indicated with an asterisks.

For 14 ASE SNPs, this is the first time genotype–expression associations are reported (referred to as “novel associations” in ▶Table 3).

Most Novel ASE SNPs (or Proxies) are in Predicted Promoters or Enhancers

ASE SNPs can affect gene expression by, for example, altering promoters, enhancers, or TF binding sites. We used HaploReg and SNP2TFBS to assess the effect of the 14 novel ASE SNPs on regulatory motifs and also to determine the chromatin state in liver tissue per haplotype block. Thirteen novel ASE SNPs (or proxies) have chromatin states indicative of promoters and/or enhancer elements in liver tissue. Eleven of these are predicted to abolish, create, or change the affinity of one or several TF binding sites (▶Table 3 and ▶Supplementary Table S3, available in the online version).

Most ASE SNPs (or Proxies) are Associated with Circulating Protein Levels or Other Relevant Traits

Given the range of processes that affect protein abundance, a SNP influencing mRNA levels may not necessarily influence protein levels (correlated with disease status). To assess a possible role of ASE SNPs in regulating levels of the respective circulating protein levels, we intersected the ASE SNPs identified here with GWAS trait-associated SNPs. A total of 29 ASE SNPs (or proxies) were previously associated with the respective protein levels (e.g., C4BPB, fibrinogen, FVII, FXII, FXIII B, α 1-antitrypsin [A1AT], and TAFI; ▶Table 2), including 5 novel ASE SNPs (▶Table 3; for a complete list see ▶Supplementary Table S4, available in the online version). Using our published GWAS on intact TAFI and TAFI-AP,⁵ we determined that the minor allele of rs2277440 (a “novel” ASE SNP in *CPB2*) is associated with lower intact TAFI and TAFI-AP levels ($p = 0.0019$ and $p = 0.038$, respectively). However, no

association was observed for rs17843995 (a validated SNP in *CPB2*; $p > 0.05$). In addition, we investigated whether rs6092 (in *SERPINE1*, without published GWAS association to protein levels), associated with circulating plasma levels of PAI-1 in $n = 600$ controls from our case–control study cohort SAHLSIS. Individuals homozygous for the minor allele of rs6092 exhibited significantly lower PAI-1 protein levels as compared with individuals homozygous for the major allele (n , median [interquartile range] minor vs. major allele: $n = 4$, 23 [20–38] ng/mL vs. $n = 485$, 40 [24–61] ng/mL; $p = 0.026$). Thus, 58% ($n = 31/53$) of ASE SNPs are associated to levels of the respective proteins.

We next assessed associations of ASE SNPs with relevant disease and nondisease phenotypes. A total of 17 ASE SNPs were associated with complex diseases (e.g., venous thromboembolism, coronary artery disease [CAD], and stroke) and other related phenotypes (e.g., activated partial thromboplastin time, aortic valve/mitral annular calcification, hypertension, self-reported high cholesterol; and plasma levels of lipoprotein A, coagulation FX, FXIII A, and Z-dependent protease inhibitor [ZPI]), including 5 novel ASE SNPs (▶Tables 2 and 3).

Verification and Replication of ASE SNPs

We used rhAMP real-time PCR as an alternative method to verify 2 ASE SNPs identified in the initial set of 17 liver samples: rs7337140 in *CPB2* and rs2227424 in *FGB*. The heterozygous individuals displaying ASE using the custom RNAseq approach also displayed ASE using rhAMP assays (rs7337140, $n = 1$, ASE ratio 0.67; rs2227424, $n = 6$, mean ASE ratio $0.81 \pm$ standard deviation, 0.08); the negative control (rs12937244 in *SERPINA1*) showed no sign of ASE in any sample ($n = 10$, mean ASE ratio 0.5 ± 0.08). Additionally, these findings were replicated in a further set of 48

Table 2 Previously validated ASE SNPs (previously reported liver eQTLs) and novel-to-liver SNPs (previously reported eQTLs in other tissue)

dbSNP	Gene	Chr	Position (Hg19)	SNP location	Nr het	Mean allelic ratio	SE of mean ratio	eQTL validation	Respective protein level association	Other GWAS trait association
rs3813948	C4BPB	1	207269858	Intron	1	0.30	–	Validated	C4BPB	MPV
rs1087	CPB2	13	46627439	3' UTR	7	0.24	0.04	Novel to liver	TAFI	–
rs17843995	CPB2	13	46672921	Intron	6	0.08	0.02	Validated	–	–
rs2296642	CPB2	13	46656669	Synonymous	2	0.33	0.04	Novel to liver	TAFI	–
rs3742264	CPB2	13	46648094	Missense	7	0.73	0.03	Validated	TAFI	–
rs7337140 ^b	CPB2	13	46641481	Synonymous	2	0.72	0.01	Novel to liver	TAFI	–
rs940	CPB2	13	46627480	3' UTR	4	0.28	0.03	Novel to liver	TAFI-AP	–
rs3733403	F11	4	187187135	5' UTR start gained	3	0.39	0.03	Novel to liver	–	–
rs1801020	F12	5	176836532	5' UTR	7	0.64	0.01	Validated	FXII	APTT, CAD
rs5990	F13B	1	197008597	Intron	4	0.05	0.01	Novel to liver	FXIIIB	FXIIIA
rs5998	F13B	1	197009798	Synonymous	9	0.39	0.01	Novel to liver	FXIIIB	–
rs491098	F7	13	113769346	Intron	2	0.76	0.02	Validated	FVII, FVIIa	PT, FX
rs2059502	FGB	4	155493319	3' UTR	6	0.83	0.01	Novel to liver	Fibrinogen	–
rs2059503	FGB	4	155493419	3' UTR	5	0.78	0.01	Novel to liver	Fibrinogen	–
rs2227395	FGB	4	155484511	Intron	3	0.60	0.07	Validated	Fibrinogen	–
rs2227401	FGB	4	155486381	Intron	6	0.85	0.02	Validated	Fibrinogen	–
rs2227411	FGB	4	155489085	Intron	4	0.60	0.07	Novel to liver	Fibrinogen	–
rs2227423	FGB	4	155492249	3' UTR	6	0.74	0.01	Novel to liver	Fibrinogen	–
rs2227424 ^b	FGB	4	155492720	3' UTR	6	0.78	0.01	Novel to liver	Fibrinogen	–
rs2227426	FGB	4	155493171	3' UTR	6	0.76	0.01	Novel to liver	Fibrinogen	–
rs4681	FGB	4	155490832	Synonymous	6	0.61	0.02	Novel to liver	Fibrinogen	–
rs6056	FGB	4	155488821	Synonymous	6	0.82	0.03	Novel to liver	Fibrinogen	–
rs2066865	FGG	4	155525276	Downstream	4	0.67	0.01	Novel to liver	Fibrinogen	Stroke, VT, PE, phlebitis, and thrombophlebitis
rs4253331	KLKB1	4	187179135	Intron	3	0.23	0.05	Novel to liver	–	Treatment with antihypertensive
rs1043334	LMAN1	18	56996526	3' UTR	4	0.38	0.02	Novel to liver	–	–
rs11060	PLG	6	161173946	Synonymous	4	0.73	0.02	Novel to liver	–	Hypertension
rs13231	PLG	6	161139857	Synonymous	5	0.62	0.02	Novel to liver	–	LpA, CAD, s-r MI, s-r angina, s-r high cholesterol; aortic valve and mitral annular calcification
rs1130656	PLG	6	161139480	Synonymous	3	0.67	0.09	Novel to liver	–	LpA
rs4252086	PLG	6	161132725	3' UTR	1	0.12	–	Novel to liver	–	LpA, CAD, CIHD, aortic valve, and mitral annular calcium
rs9681204	PROS1	3	93592569	3' UTR	5	0.22	0.01	Novel to liver	–	–
rs2273971	PROZ	13	113812962	Upstream	3	0.68	0.03	Novel to liver	–	–
rs1028580	SERPINA1	14	94849882	Intron	4	0.23	0.09	Novel to liver	–	–
rs1243164	SERPINA1	14	94844470	3' UTR	8	0.37	0.02	Novel to liver	A1AT	ZPI
rs1243166	SERPINA1	14	94843818	3' UTR	8	0.27	0.02	Novel to liver	–	–
rs6647	SERPINA1	14	94847415	Missense	6	0.84	0.03	Novel to liver	A1AT	–
rs709932	SERPINA1	14	94849201	Missense	2	0.69	0.01	Novel to liver	A1AT	ZPI
rs2069970	SERPINA5	14	95053286	Intron	1	0.10	–	Novel to liver	–	–

(Continued)

Table 2 (Continued)

dbSNP	Gene	Chr	Position (Hg19)	SNP location	Nr het	Mean allelic ratio	SE of mean ratio	eQTL validation	Respective protein level association	Other GWAS trait association
rs6107 ^a	SERPINA5	14	95057270	Intron	7	0.70	0.07	Novel to liver	–	–
rs5877	SERPINC1	1	173878862	Synonymous	9	0.64	0.01	Novel to liver	–	–

Abbreviations: APTT, activated partial thromboplastin time; ASE, allele-specific expression; A1AT, α 1-antitrypsin (SERPINA1); C4BPB, complement C4b binding protein; CAD, coronary artery disease; CIHD, chronic ischemic heart disease; dbSNP, Database of Single Nucleotide Polymorphisms; eQTL, expression quantitative trait loci; FVII, factor VII; FVIIa, FVII activity; FX, factor X; FXII, factor XII; FXIII, factor XIII; FXIIIa, FXIII A subunit; FXIIIb, FXIII B subunit; FXIIIa, FXIII activity; GWAS, genome-wide association study; LpA, lipoprotein A; MI, myocardial infarction; MPV, mean platelet volume; PE, pulmonary embolism; PT, prothrombin time; s-r, self-reported; SNP, single nucleotide polymorphism; TAFI, thrombin activatable fibrinolysis inhibitor; TAFI-AP, TAFI activation peptide; UTR, untranslated region; VT, venous thrombosis; ZPI, Z-dependent protease inhibitor (SERPINA10).

Note: Chr, chromosome number; Nr het, number of heterozygotes based on $n = 17$ genotype and expression profiled samples; Mean allelic ratio, mean allelic imbalance averaged among all heterozygous individuals (standardized so alternative allele is the minor allele); SE of mean ratio, standard error of mean allelic ratio from all heterozygous individuals.

^aIndicates bidirectional allelic preference: $n = 1$ individual displayed an allelic imbalance < 0.35 and $n = 6$ displayed an allelic imbalance > 0.65 .

^bSNPs selected for rhAMP validation and replication.

liver samples, with individuals heterozygous for rs7337140 having an average allelic ratio of 0.73 (± 0.08 , $n = 6$) and rs2227424 having an average allelic ratio of 0.79 (± 0.09 , $n = 10$), in line with the custom RNAseq results.

Discussion

In this study, we identified widespread ASE of hemostatic genes in human liver, and assessed the novelty, and importance, of these SNPs for circulating levels of the respective protein as well as relevant disease and nondisease phenotypes via public eQTL and GWAS data sets. Many of these SNPs have previously been associated with levels of the respective transcripts ($n = 39/53$) and/or circulating plasma levels of the respective proteins ($n = 29/53$), and others with thrombotic or bleeding disorders or related traits ($n = 17/53$). More importantly, this differential allelic approach using targeted-sequencing also lead to the identification of 14 novel genotype–expression associations. To our knowledge, this is the most comprehensive *cis*-regulatory data set for hemostatic genes in liver to date. Thus, we believe this study will be a useful resource for designing experimental studies to decipher the underlying mechanisms regulating expression of these genes that may play a direct role in conferring thrombotic and bleeding risk.

Using ASE analysis, we identified 14 novel putative regulatory SNPs (termed “novel ASE SNPs”), not previously reported in public eQTL databases across any tissues (i.e., this is the first ever reported genotype–expression association). Public GWAS data were used to identify 8 novel ASE SNPs with robust associations to relevant traits. These include 5 ASE SNPs associated with circulating levels of their respective proteins (4 SNPs in the genes encoding fibrinogen^{29–31} and one in the gene encoding FXIII B³²; **Table 3** and **Supplementary Table S4**, available in the online version). We further demonstrated that an additional novel ASE SNP (in *CPB2*) was associated with circulating TAFI levels. Two of the aforementioned and three others were associated with thrombotic diseases

(e.g., venous thromboembolism³³ and CAD³⁴) and other related phenotypes (e.g., activated partial thromboplastin time³⁵; aortic valve/mitral annular calcification³⁶; and systolic blood pressure, UK Biobank; **Table 3**; **Supplementary Table S4**, available in the online version). Of these nine SNPs, eight (or proxies) have chromatin states indicative of promoter and/or enhancers in liver and are predicted to alter one or several TF binding sites. Thus, all nine novel ASE SNPs have features indicative of a plausible biologically relevant, and consequently a plausible clinically relevant, function.

For the remaining five novel ASE SNPs, an extensive search of the literature lead to the identification of two SNPs previously shown to associate with quantitative, thrombotic, or bleeding traits. For the first, in the gene encoding PAI-1 (rs6092, Ala15Thr), the minor allele was reported to associate with decreased PAI-1 antigen levels and activity and bleeding after trauma and surgery in a case report³⁷ and have a protective role in myocardial infarction.³⁸ In line with this, using controls from SAHLIS, we showed that individuals homozygous for the minor allele have significantly lower PAI-1 antigen levels compared with individuals homozygous for the major allele. As this SNP is within the coding region, it may alter the function and/or stability of the PAI-1 protein itself. However, this SNP also has a chromatin state indicative of an enhancer in liver, and is predicted to alter four TF binding sites (i.e., BDP1, HNF4, NF-1, and RXRA, according to HaploReg). Thus, in addition to having a direct affect on PAI-1 protein function, this SNP could have a role in *SERPINE1* gene expression, a hypothesis that requires experimental validation.

The second SNP, in the gene encoding ZPI (also known as SERPINA10; rs61754487, Trp303x), is a rare stop-gained (nonsense) mutation associated with recurrent spontaneous miscarriage.³⁹ This SNP would be expected to result in the loss of ZPI activity and consequently a thrombotic phenotype. It has been reported to be linked to venous thrombosis in a small candidate study (where $n = 8/250$ cases carried this mutation vs. $n = 0/250$ controls⁴⁰). A later meta-analysis

Table 3 Novel genotype–expression associations (eQTLs not previously reported in any tissue)

dbSNP	Gene	Chr	Position (Hg19)	SNP location	Nr het	Mean allelic ratio	SE of mean ratio	Respective protein level association	Other GWAS trait association	HaploReg promoter	HaploReg enhancer	Predicted regulatory motif changed
rs2277440 ^d	CPB2	13	46638826	Synonymous	2	0.73	0.06	TAFI, TAFI-AP ^d	–	–	–	–
rs3212994	F10	13	113777130	5' UTR start gained	1	0.35	–	–	–	Yes	Yes	Yes ^b
rs5993	F13B	1	197030829	Intron	6	0.40	0.03	FXIIIB, FXIIIa	FXIIIa, SBP	Yes	Yes	Yes ^{b,c}
rs2070018	FGA	4	15508627	Intron	5	0.64	0.01	Fibrinogen	–	Yes	Yes	Yes ^{b,c}
rs1044291	FGB	4	155493352	3' UTR	5	0.35	0.05	Fibrinogen	–	Yes	Yes	Yes ^{b,c}
rs2227412	FGB	4	155489095	Intron	2	0.27	0.02	Fibrinogen	–	Yes	Yes	Yes ^{b,c}
rs1049636	FGG	4	155525970	3' UTR	8	0.38	0.02	Fibrinogen	VT	Yes	Yes	Yes ^{b,c}
rs3856930	KNG1	3	186458322	Intron	1	0.84	–	–	APTT	–	Yes	Yes ^{b,c}
rs5030100	KNG1	3	186440147	Intron	1	0.75	–	–	–	–	Yes	Yes ^b
rs2859879	PLG	6	161160083	Intron	7	0.27	0.03	–	CAD	–	Yes ^a	Yes ^b
rs4063598	PLG	6	161122931	Upstream	1	0.89	–	–	Aortic valve and mitral annular calcification	Yes	Yes	Yes ^{b,c}
rs201774333	SERPINA1	14	94844947	Missense	2	0.79	0.03	–	–	–	Yes	–
rs61754487	SERPINA10	14	94754643	Stop gained	1	0.87	–	–	–	–	Yes	–
rs6092 ^e	SERPINE1	7	100771717	Missense	2	0.60	0.06	PAI-1 ^e	–	–	Yes ^a	Yes

Abbreviations: APTT, activated partial thromboplastin time; CAD, coronary artery disease; dbSNP, Database of Single Nucleotide Polymorphisms; eQTL, expression quantitative trait loci; FXIIIA, factor XIII A subunit; FXIIIB, factor XIII B subunit; FXIIIa, FXIII activity; GWAS, genome-wide association study; IQR, interquartile range; SBP, systolic blood pressure; SNP, single nucleotide polymorphism; TAFI, thrombin activatable fibrinolysis inhibitor; TAFI-AP, TAFI activation peptide; UTR, untranslated region; VT, venous thrombosis.

Note: Chr, chromosome number; Nr het, number of heterozygotes based on $n = 17$ genotype and expression profiled samples; Mean allelic ratio, mean allelic imbalance averaged among all heterozygous individuals (standardized so alternative allele is the minor allele); SE of mean ratio, standard error of mean allelic ratio from all heterozygous individuals.

^aH3K4me3 peaks were used to define enhancer, rather than the default 15-state core ChromHMM model.

^bPredicted regulatory motif changed based on data from HaploReg.

^cSNP predicted to abolish, create, or change the affinity of transcription factor binding sites, based on SNP2TFBS.

^dThe minor allele of rs2277440 is associated with lower intact TAFI and TAFI-AP plasma levels, $p = 0.0019$ and $p = 0.038$, respectively.⁵

^ePAI-1 plasma levels were lower in control individuals from SAHLSIS homozygous for the minor allele of rs6092 as compared with major allele (n , median [IQR] for minor vs. major allele: $n = 4$, 23 [20–38] ng/mL vs. $n = 485$, 40 [24–61] ng/mL; $p = 0.026$).

did not confirm this association, but it is of note that this rare variant was present in only $n = 14/2,311$ venous thromboembolism cases and $n = 12/2,821$ controls.⁴¹ This SNP is not predicted to alter any regulatory motifs. However, rather than having identified a potential *cis*-regulatory SNP, we are likely observing ASE resulting from nonsense-mediated decay, a conserved posttranslational pathway whereby transcripts bearing aberrant termination codons are selectively identified and eliminated.⁴²

The remaining three novel ASE SNPs have no previous mention in the literature, as far as we are aware. These were in genes encoding coagulation FX (rs3212994), with a chromatin state indicative of a promoter in liver tissue and predicted to alter two regulatory motifs (BHLHE40 and NRSF/REST); kininogen 1 (rs5030100) with an enhancer chromatin signature in liver and predicted to alter the binding site of three TFs (Ik-3, NF-kappaB, STAT); and A1AT (also known as SERPINA10, rs201774333) with an enhancer chromatin signature.

Taken together, most novel ASE SNPs identified in our study have features indicative of a plausible biologically relevant function. Our data provide another piece to the puzzle for how genetic variants affect hemostatic protein levels and risk for thrombotic or bleeding disorders. Above, we provide several examples of SNPs that associated with mRNA levels, protein levels, and disease phenotypes, which have relevant chromatin signatures in the relevant tissue (liver) and are predicted to alter TF motifs. All of these would be good candidates for functional studies. The same logic can be applied to many of the other ASE SNPs we identified. In the future, we envisage that this data set could be used by the hemostatic research community. For example, if a hemostatic SNP is found to associate with a thrombotic or bleeding trait, this data can be used to see if the SNP displays ASE in liver. In this way, it could be a stepping stone for designing appropriate experimental studies.

As outlined in the introduction, public eQTL data currently have low coverage of hemostatic genes in liver tissue and in line with this only 7 of the 53 ASE SNPs identified here had previously been associated with expression of the respective genes in liver (liver eQTLs, termed “previously validated”). Interestingly, the majority of these validated SNPs have previously been shown to associate with plasma levels of the respective proteins and/or thrombotic or bleeding traits. Three examples include: (1) a 5' untranslated region variant in *F12* (rs1801020) associated with circulating FXII concentration, delayed activated partial thromboplastin time,⁴³ in-hospital death after alteplase,⁴⁴ CAD,⁴⁵ ischemic stroke,⁴⁶ and venous thrombosis during first pregnancy⁴⁷; (2) an intron variant in *F7* (rs491098), associated to FVII levels and prothrombin time^{32,35}; and (3) an intron variant in *FGB* (rs2227401) associated with fibrinogen levels.³¹ These findings underscore the validity of our ASE approach.

For the most part, for a particular ASE SNP, the preferably expressed allele was consistent between individuals although the effect size (allelic ratio) varied. However, one SNP exhibited bidirectional allelic preference (rs6107, an intronic variant in *SERPINA5*). This could result from allelic heterogeneity of one or more *cis*-acting regulatory polymorphisms present in cod-

ing, intronic, and/or other noncoding regulatory sequences⁴⁸ or from epigenetic factors such as DNA methylation.⁴⁹ Alternatively, the SNP used in the ASE analysis could be in LD with another SNP that is responsible for the differential expression of allelic transcripts, but that does not show unidirectional ASE due to incomplete LD with the marker SNP.

ASE has an advantage over eQTL analyses because it uses a within-sample control (the other allele) to query effects of SNPs on expression levels. In doing so, the ASE approach eliminates sources of error between individuals that alter the overall expression of that gene (e.g., *trans*-, environmental or experimental effects). Thus, the ASE approach can detect SNPs that are masked by, for example, local *trans* effects and feedback mechanisms in eQTL studies.¹⁵ In addition to this methodological strength, we used sequencing which is largely immune to the SNP-in-probe effect known to inflate false positives in classical ASE analysis using microarrays. To further overcome bias in our study, we used N-masked individualized genomes based on gDNAseq data to prevent alignment bias. Additionally, the custom target sequence design used here provides much deeper coverage across the genes of interest compared with whole transcriptome-seq and avoids the disadvantage of off-target noise and enables determination of each SNP in all individual samples. Finally, we used an independent method (rhAMP real-time PCR) and liver samples to replicate selected findings proving that this method is reliable. Despite these strengths, there are some limitations which deserve mention. The targeted design we chose was limited to 5 kb upstream and 0.5 kb downstream of each gene, which means we could have missed important regulatory information outside of our target area. For the assessment of novelty of identified ASE SNPs, we used an LD cut-off of $r^2 > 0.8$. However, we cannot rule out that some variants we classify as “novel” are in fact in LD with known eQTL SNPs with very different allele frequencies (i.e., low r^2 but high D'). Although the ASE approach is powerful, our study is based on just 17 samples. Therefore, novel findings should be replicated in an independent data set. Further, we acknowledge that extensive functional validation is required to better understand the association between ASE SNPs and quantitative traits/disease associations. Additionally, it should be emphasized that the results were generated in European ancestry populations, and our findings may not be generalizable to populations of different ethnicity.

In summary, using targeted DNA- and RNAseq, we have shown widespread ASE in hemostatic genes in human liver (53 SNPs in 21/35 genes). We validated 74% ($n = 39/53$) of these SNPs in public eQTL databases and also identified 14 novel potential regulatory SNPs. Moreover, approximately 55% of ASE SNPs were previously associated with altered circulating levels of their respective proteins. Of the novel ASE SNPs, the majority had chromatin signatures consistent with promoter and/or enhancer in liver and were predicted to alter TF binding sites. Taken together, we identified many SNPs with plausible functional mechanisms that may play a role in conferring thrombotic or bleeding risk. To our knowledge, this is the most comprehensive *cis*-regulatory resource for hemostatic genes in liver to date. We believe that this data set will be a

useful resource for designing experimental studies to unravel the underlying mechanisms regulating expression of these important genes and may enhance the identification of therapeutic targets for thrombotic and bleeding disorders.

What is known about this topic?

- The liver plays a major role in producing proteins that are secreted into the blood, including many factors involved in hemostasis. Understanding the genetic basis underlying hepatic hemostatic gene expression variability may reveal genetic variants that are important for thrombotic and/or bleeding disorders.
- Publically available expression quantitative trait loci (eQTL) data sets currently have low coverage of eQTLs in hemostatic genes in human liver tissue. Allele-specific expression (ASE) analysis is an alternative way to quantify *cis*-regulatory variation, and can be performed in a targeted manner to focus on a particular set of genes.
- To the best of our knowledge, no studies have analyzed ASE in hemostatic genes in human liver.

What does this paper add?

- Using targeted DNA- and RNA-sequencing, our study identified widespread ASE in hemostatic genes (53 ASE SNPs corresponding to 21 hemostatic genes). Of these, 7 ASE SNPs were previously validated as eQTLs for the relevant genes in liver tissue using public data sets and 32 were eQTLs in other cell types/tissues.
- A total of 29 ASE SNPs were previously associated with the respective plasma protein levels in genome-wide association studies (GWAS), and 17 ASE SNPs to other relevant traits including venous thromboembolism, coronary artery disease, and stroke.
- More importantly, this differential allelic approach also leads to the identification of 14 novel genotype-expression associations. Thus, to our knowledge this is the most comprehensive *cis*-regulatory resource for hemostatic genes in liver to date.

Authors' Contributions

T.M.S. and C.J. conceived the research design of the present study. C.J. was responsible for sample contribution, M.O.L. and S.K. isolated gDNA and RNA and M.O.L. prepared sequencing libraries. M.O.L., L.H., and M.D.L. performed statistical analyses. M.O.L., L.H., S.K., and M.D.L. drafted the figures. M.O.L., C.J., and T.M.S. interpreted the data. M.O.L. and T.M.S. drafted the manuscript. L.H., M.D.L., and C.J. intellectually reviewed the manuscript. All authors contributed to the last revision process and approved the version to be published.

Funding

This study was supported by the Swedish Heart and Lung Foundation (20160316), the Swedish Research Council

(2018-02543), the Swedish Stroke Association, the Swedish state under the agreement between the Swedish government and the county councils (the ALF-agreement, ALFGBG-42), the Swedish Foundation for Strategic Research (RIF14-0081), the Rune and Ulla Amlövs Foundation for Neurologic Research, the John and Brit Wennerström Foundation for Neurologic Research, the Marcus Borgströms Foundation for Neurologic Research, and the Nilsson-Ehle Endowments.

Conflict of Interest

None declared.

Acknowledgments

The authors thank Prof. Olausson and Dr. Rizell, the Transplant Centre, Sahlgrenska University Hospital, and other "Liver Team members" for technical and administrative assistance in providing the liver specimens. We thank the sequencing and genotyping service provided by Genomics Core Facility at the University of Gothenburg, Gothenburg, Sweden; Science for Life Laboratory (SciLifeLab), Stockholm, Sweden; and TATAA Biocenter AB, Gothenburg, Sweden. The computations were performed using resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). The authors gratefully acknowledge support from the Bioinformatics Long-term Support, SciLifeLab (WABI, Stockholm and Uppsala Sweden) for RNA sequencing data analysis.

References

- 1 de Vries PS, Chasman DI, Sabater-Lleal M, et al. A meta-analysis of 120 246 individuals identifies 18 new loci for fibrinogen concentration. *Hum Mol Genet* 2016;25(02):358–370
- 2 Smith NL, Chen MH, Dehghan A, et al; Wellcome Trust Case Control Consortium. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: the CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation* 2010;121(12):1382–1392
- 3 van Loon J, Dehghan A, Weihong T, et al. Genome-wide association studies identify genetic loci for low von Willebrand factor levels. *Eur J Hum Genet* 2016;24(07):1035–1040
- 4 Huang J, Huffman JE, Yamakuchi M, et al; Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) Consortium Neurology Working Group; CARDIoGRAM Consortium; CHARGE Consortium Hemostatic Factor Working Group. Genome-wide association study for circulating tissue plasminogen activator levels and functional follow-up implicates endothelial STXBP5 and STX2. *Arterioscler Thromb Vasc Biol* 2014;34(05):1093–1101
- 5 Stanne TM, Olsson M, Lorentzen E, et al. A genome-wide study of common and rare genetic variants associated with circulating thrombin activatable fibrinolysis inhibitor. *Thromb Haemost* 2018;118(02):298–308
- 6 Dehghan A, Bis JC, White CC, et al. Genome-wide association study for incident myocardial infarction and coronary heart disease in prospective cohort studies: the CHARGE Consortium. *PLoS One* 2016;11(03):e0144997
- 7 Klarin D, Emdin CA, Natarajan P, Conrad MF, Kathiresan S; INVENT Consortium. Genetic analysis of venous thromboembolism in UK Biobank identifies the ZFPM2 locus and implicates obesity as a causal risk factor. *Circ Cardiovasc Genet* 2017;10(02):10

- 8 Malik R, Chauhan G, Traylor M, et al; AFGen Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium; International Genomics of Blood Pressure (iGEN-BP) Consortium; INVENT Consortium; STARNET; BioBank Japan Cooperative Hospital Group; COMPASS Consortium; EPIC-CVD Consortium; EPIC-InterAct Consortium; International Stroke Genetics Consortium (ISGC); METASTROKE Consortium; Neurology Working Group of the CHARGE Consortium; NINDS Stroke Genetics Network (SiGN); UK Young Lacunar DNA Study; MEGASTROKE Consortium; MEGASTROKE Consortium. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 2018; 50(04):524–537
- 9 Falcone GJ, Woo D. Genetics of spontaneous intracerebral hemorrhage. *Stroke* 2017;48(12):3420–3424
- 10 Konkle BA, Johnsen JM, Wheeler M, Watson C, Skinner M, Pierce GF; My Life Our Future programme. Genotypes, phenotypes and whole genome sequence: approaches from the My Life Our Future haemophilia project. *Haemophilia* 2018;24(Suppl 6):87–94
- 11 Consortium G; GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348(6235):648–660
- 12 Innocenti F, Cooper GM, Stanaway IB, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet* 2011;7(05):e1002078
- 13 Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 2007;39(02):226–231
- 14 Pickrell JK, Marioni JC, Pai AA, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 2010;464(7289):768–772
- 15 Campino S, Forton J, Raj S, et al. Validating discovered Cis-acting regulatory genetic variants: application of an allele specific expression approach to HapMap populations. *PLoS One* 2008;3(12):e4105
- 16 Mercer TR, Clark MB, Crawford J, et al. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat Protoc* 2014;9(05):989–1009
- 17 Dimas AS, Deutsch S, Stranger BE, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 2009;325(5945):1246–1250
- 18 Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol* 2015;16:195
- 19 Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* 2016;44(D1):D877–D881
- 20 Kundaje A, Meuleman W, Ernst J, et al; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518(7539):317–330
- 21 Kumar S, Ambrosini G, Bucher P. SNP2TFBS - a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res* 2017;45(D1):D139–D144
- 22 Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 2016;32(20):3207–3209
- 23 Battle A, Brown CD, Engelhardt BE, Montgomery SB; GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC) –Analysis Working Group; Statistical Methods groups –Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site –NDRI; Biospecimen Collection Source Site –RPCI; Biospecimen Core Resource –VARI; Brain Bank Repository –University of Miami Brain Endowment Bank; Leidos Biomedical –Project Management; ELSI Study; Genome Browser Data Integration & Visualization –EBI; Genome Browser Data Integration & Visualization –UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis & Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group. Genetic effects on gene expression across human tissues. *Nature* 2017; 550(7675):204–213
- 24 Eicher JD, Landowski C, Stackhouse B, 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
- 25 Franzén O, Ermel R, Cohain A, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* 2016;353(6301):827–830
- 26 Jood K, Ladenvall C, Rosengren A, Blomstrand C, Jern C. Family history in ischemic stroke before 70 years of age: the Sahlgrenska Academy Study on Ischemic Stroke. *Stroke* 2005;36(07):1383–1387
- 27 Jood K, Ladenvall P, Tjärnlund-Wolf A, et al. Fibrinolytic gene polymorphism and ischemic stroke. *Stroke* 2005;36(10):2077–2081
- 28 Söderholm M, Almgren P, Jood K, et al. Exome array analysis of ischaemic stroke: results from a southern Swedish study. *Eur J Neurol* 2016;23(12):1722–1728
- 29 Danik JS, Paré G, Chasman DI, et al. Novel loci, including those related to Crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of fibrinogen in 17 686 women: the Women's Genome Health Study. *Circ Cardiovasc Genet* 2009;2(02):134–141
- 30 Lovely RS, Yang Q, Massaro JM, et al. Assessment of genetic determinants of the association of γ' fibrinogen in relation to cardiovascular disease. *Arterioscler Thromb Vasc Biol* 2011;31(10):2345–2352
- 31 Wassel CL, Lange LA, Keating BJ, et al. Association of genomic loci from a cardiovascular gene SNP array with fibrinogen levels in European Americans and African-Americans from six cohort studies: the Candidate Gene Association Resource (CARE). *Blood* 2011;117(01):268–275
- 32 Williams FM, Carter AM, Hysi PG, et al; EuroCLOT Investigators; Wellcome Trust Case Control Consortium 2; MOnica Risk, Genetics, Archiving and Monograph; MetaStroke; International Stroke Genetics Consortium. Ischemic stroke is associated with the ABO locus: the EuroCLOT study. *Ann Neurol* 2013;73(01):16–31
- 33 Tang W, Teichert M, Chasman DI, et al. A genome-wide association study for venous thromboembolism: the extended Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. *Genet Epidemiol* 2013;37(05):512–521
- 34 van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res* 2018;122(03):433–443
- 35 Tang W, Schwienbacher C, Lopez LM, et al. Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. *Am J Hum Genet* 2012;91(01):152–162
- 36 Thanassoulis G, Campbell CY, Owens DS, et al; CHARGE Extra-coronary Calcium Working Group. Genetic associations with valvular calcification and aortic stenosis. *N Engl J Med* 2013; 368(06):503–512
- 37 Zhang ZY, Wang ZY, Dong NZ, Bai X, Zhang W, Ruan CG. A case of deficiency of plasma plasminogen activator inhibitor-1 related to Ala15Thr mutation in its signal peptide. *Blood Coagul Fibrinolysis* 2005;16(01):79–84
- 38 Morange PE, Saut N, Alessi MC, et al. Association of plasminogen activator inhibitor (PAI)-1 (SERPINE1) SNPs with myocardial infarction, plasma PAI-1, and metabolic parameters: the HIFMECH study. *Arterioscler Thromb Vasc Biol* 2007;27(10):2250–2257
- 39 AlShaikh FS, Finan RR, Almawi AW, Mustafa FE, Almawi WY. Association of the R67X and W303X non-sense polymorphisms in the protein Z-dependent protease inhibitor gene with idiopathic recurrent miscarriage. *Mol Hum Reprod* 2012;18(03):156–160

- 40 Van de Water N, Tan T, Ashton F, et al. Mutations within the protein Z-dependent protease inhibitor gene are associated with venous thromboembolic disease: a new form of thrombophilia. *Br J Haematol* 2004;127(02):190–194
- 41 Dentali F, Gianni M, Lussana F, Squizzato A, Cattaneo M, Ageno W. Polymorphisms of the Z protein protease inhibitor and risk of venous thromboembolism: a meta-analysis. *Br J Haematol* 2008;143(02):284–287
- 42 Kukurba KR, Zhang R, Li X, et al. Allelic expression of deleterious protein-coding variants across human tissues. *PLoS Genet* 2014;10(05):e1004304
- 43 Roldán V, Corral J, Marín F, Pineda J, Vicente V, González-Conejero R. Synergistic association between hypercholesterolemia and the C46T factor XII polymorphism for developing premature myocardial infarction. *Thromb Haemost* 2005;94(06):1294–1299
- 44 del Río-Espínola A, Fernández-Cadenas I, Giralt D, et al; GRECOS Investigators. A predictive clinical-genetic model of tissue plasminogen activator response in acute ischemic stroke. *Ann Neurol* 2012;72(05):716–729
- 45 Santamaría A, Martínez-Rubio A, Mateo J, Tirado I, Soria JM, Fontcuberta J. Homozygosity of the T allele of the 46 C->T polymorphism in the F12 gene is a risk factor for acute coronary artery disease in the Spanish population. *Haematologica* 2004;89(07):878–879
- 46 Santamaría A, Mateo J, Tirado I, et al. Homozygosity of the T allele of the 46 C->T polymorphism in the F12 gene is a risk factor for ischemic stroke in the Spanish population. *Stroke* 2004;35(08):1795–1799
- 47 Cochery-Nouvellon E, Mercier E, Lissalde-Lavigne G, et al. Homozygosity for the C46T polymorphism of the F12 gene is a risk factor for venous thrombosis during the first pregnancy. *J Thromb Haemost* 2007;5(04):700–707
- 48 Hardison RC, Oeltjen J, Miller W. Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res* 1997;7(10):959–966
- 49 Schalkwyk LC, Meaburn EL, Smith R, et al. Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet* 2010;86(02):196–212