A Comprehensive Sequencing-Based Analysis of Allelic Methylation Patterns in Hemostatic Genes in Human Liver

N.B. When citing this work, cite the original published paper.

(article starts on next page)

THIEME OPEN ACCESS

A Comprehensive Sequencing-Based Analysis of Allelic Methylation Patterns in Hemostatic Genes in Human Liver

Martina Olsson Lindvall[1]   Marcela Davila Lopez[2]   Sofia Klasson[1]   Lena Hansson[3,4]   Staffan Nilsson[1,5]   Tara M. Stanne[1,*]   Christina Jern[1,6,*]

[1] Department of Laboratory Medicine, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
[2] Bioinformatics Core Facility, University of Gothenburg, Gothenburg, Sweden
[3] Science for Life Laboratories (SciLifeLab), Stockholm, Sweden
[4] Novo Nordisk, Oxford, United Kingdom
[5] Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden
[6] Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital, Gothenburg, Sweden

Address for correspondence Martina Olsson Lindvall, MSc, Institute of Biomedicine, Sahlgrenska Academy at the University of Gothenburg, Box 445, SE-405 30 Gothenburg, Sweden (e-mail: martina.olsson@gu.se).

Abstract

Characterizing the relationship between genetic, epigenetic (e.g., deoxyribonucleic acid [DNA] methylation), and transcript variation could provide insights into mechanisms regulating hemostasis and potentially identify new drug targets. Several hemostatic factors are synthesized in the liver, yet high-resolution DNA methylation data from human liver tissue is currently lacking for these genes. Single-nucleotide polymorphisms (SNPs) can influence DNA methylation in *cis* which can affect gene expression. This can be analyzed through allele-specific methylation (ASM) experiments. We performed targeted genomic DNA- and bisulfite-sequencing of 35 hemostatic genes in human liver samples for SNP and DNA methylation analysis, respectively, and integrated the data for ASM determination. ASM-associated SNPs (ASM-SNPs) were tested for association to gene expression in liver using in-house generated ribonucleic acid-sequencing data. We then assessed whether ASM-SNPs associated with gene expression, plasma proteins, or other traits relevant for hemostasis using publicly available data. We identified 112 candidate ASM-SNPs. Of these, 68% were associated with expression of their respective genes in human liver or in other human tissues and 54% were associated with the respective plasma protein levels, activity, or other relevant hemostatic genome-wide association study traits such as venous thromboembolism, coronary artery disease, stroke, and warfarin dose maintenance. Our study provides the first detailed map of the DNA methylation landscape and ASM analysis of hemostatic genes in human liver tissue, and suggests that methylation regulated by genetic variants in *cis* may provide a mechanistic link between noncoding SNPs and variation observed in circulating hemostatic proteins, prothrombotic diseases, and drug response.

Keywords
► coagulation factors
► gene regulation
► hemostasis
► single-nucleotide polymorphisms

* Authors contributed equally.

Introduction

Plasma concentrations of hemostatic proteins vary in healthy humans and several hemostatic factors are synthesized in the liver. Characterizing the complex relationship between genetic, epigenetic (e.g., deoxyribonucleic acid [DNA] methylation), and transcriptomic variation has the potential to increase understanding about the mechanisms regulating hemostasis and to inform drug development. DNA methylation profiles vary among tissues and cell types, thus analyzing the appropriate tissue is essential. Genetic variants can exert an influence on DNA methylation in *cis*,[1,2] which can then affect *cis*-regulatory gene expression.[3,4] Consequently, it has been suggested that DNA methylation could provide the link between noncoding genetic variants and phenotypic variation.

DNA methylation at cytosine–guanine dinucleotides (CpGs) is the most studied epigenetic trait and has been most extensively studied in CpG islands in the promoter of genes, where it is widely regarded as a mechanism of transcriptional repression.[5,6] More recently, however, DNA methylation within gene bodies (i.e., exons and introns) has been reported to be prevalent and to increase gene expression.[7–10] Thus, the relationship between DNA methylation and gene expression is highly complex and not fully understood.[11,12] Most methylation studies have used microarrays that cover $< 4\%$ of human CpG sites, and this percentage is mainly biased toward CpGs in promoters.[13,14] The current gold standard method to investigate DNA methylation patterns is bisulfite conversion of the DNA followed by DNA sequencing, which allows determination of methylation patterns at a single-cytosine resolution.[13,14] Despite the potential of DNA methylation studies to contribute to new knowledge on gene regulation, high-resolution DNA methylation data are currently lacking for hemostatic genes in the liver.

Methylation quantitative trait loci (mQTL) mapping studies, which link variations in genotypes to DNA methylation levels at specific CpG sites across unrelated individuals, is one approach to elucidate whether genetic variants correlate to methylation of genes. Since DNA methylation is also influenced by environmental factors (e.g., smoking, diet, and drugs), large sample sizes are required. As a result, most mQTL studies to date have been performed in easy to access tissues such as blood, which is not necessarily the tissue of relevance. Furthermore, as most mQTL studies have used microarrays, our understanding of the roles of DNA methylation within gene bodies is currently limited.[15,16]

An alternative approach to mQTL is to compare methylation levels at specific CpG sites between each of the two alleles in the same individual. This alternative approach, referred to as allele-specific methylation (ASM), standardizes environmental and other confounding variables. Therefore, this is a powerful methodology to map putative *cis*-acting single-nucleotide polymorphisms (SNPs) even when using a small number of samples,[17] which facilitates analysis of "hard to access" human tissues of relevance such as liver. The ASM approach requires genotyping of SNPs for differentiating the alleles and simultaneous measurements of DNA

methylation at individual CpG sites. Both of these can be achieved by using targeted next-generation sequencing, which allows for the unbiased detection of all CpGs, including those within the gene body.

To identify novel regulatory elements involved in the control of hemostatic gene expression in liver we performed DNA and methylation sequencing of 35 hemostatic genes. As far as we are aware, no study has yet reported DNA methylation patterns on an individual CpG level nor analyzed ASM in hemostatic genes in human liver tissue samples. Given this background, the aims of the present study are to: (1) profile methylation patterns across hemostatic genes in human liver tissue using targeted bisulfite-sequencing; (2) identify ASM-associated SNPs (ASM-SNPs) in hemostatic genes in human liver; and (3) determine whether DNA methylation regulated by genetic variation could be a mechanistic link between noncoding SNPs and variation observed in gene expression, circulating hemostatic proteins, or other relevant traits.

Methods

Sample Collection and Targeted Sequencing

Tissue collection, genomic DNA (gDNA) and messenger ribonucleic acid (mRNA) extraction, and the target sequencing (-seq) designs for gDNA and mRNA have been previously described.[18] In short, macroscopically noncirrhotic and non-tumorous human liver tissue and blood samples were collected from adult patients of European descent ($n = 19$) undergoing liver surgery at the Sahlgrenska University Hospital, Gothenburg, Sweden. The institutional review board, i.e., the Ethics Committee of the University of Gothenburg approved the collection of the human liver tissues and blood samples, and their subsequent use for the purpose of this study. Clinical characteristics of study participants are summarized in ►**Supplementary Table S1** (available in the online version). gDNA was isolated from blood for genotyping (DNA-seq). gDNA and mRNA were simultaneously isolated from the same piece of liver tissue for methylation (bisulfite-seq) and expression (mRNA-seq) analysis, respectively.

We selected 35 genes that are predominantly 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 (MSigDB, Broad Institute). Custom capture kits (Agilent, Santa Clara, California, United States) were designed for targeted sequencing. For full details on gDNA and mRNA sequencing, data processing, and analysis, please see Olsson Lindvall et al.[18]

Targeted Gene Panel for Bisulfite-Seq

A SureSelect custom capture kit (Agilent) was designed for DNA methylation analysis (bisulfite-seq) targeting the same 35 genes as for gDNA using Agilent's SureDesign tool (Agilent). Probes were designed to capture all exons, introns, and all potential upstream (≥ 5 kb) and downstream (0.5 kb) regulatory regions, resulting in a 3.3-Mb design. Detailed information of all regions included in this study is shown in ►**Supplementary Table S2** (available in the online version). gDNA samples from liver were sheared to an average size of

250 base pair (bp) and bisulfite converted with the EZ DNA methylation kit (Zymo Research, Irvine, California, United States). The bisulfite-treated DNA was then used to prepare strand-specific sequencing libraries using the custom-designed SureSelect kits (Agilent) according to the manufacturer's protocols. All sample libraries were multiplexed in one sequence lane on the Illumina HiSeq 2500 with high output mode and 100 bp paired-end reads (Illumina, San Diego, California, United States), following the manufacturer's protocols. Approximately 15 M reads were generated per sample (for coverage per gene see ►**Supplementary Fig. S1**, available in the online version). Samples from two patients were also sequenced using SureSelect Methyl-Seq target enrichment system for Illumina (a commercially available kit from Agilent which provides targeted coverage of 84 Mb of the human genome) to identify possible weaknesses or biases in our custom-designed kit (prepared as above). Assessment of the data showed no bias in DNA methylation values between the two kits, but as expected, coverage was much higher in our custom targeted-seq samples (►**Supplementary Fig. S1**, available in the online version). All raw output sequence reads were quality controlled using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After removal of low-quality bases (Phred score < 30), reads were trimmed (removal of adaptors and indexes) with Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and aligned to the human reference genome (GRCh37/hg19) with Bowtie2.[19] Deduplication and methylation extraction were performed with the Bismark pipeline (https://www.bioinformatics.babraham.ac.uk/projects/bismark/) using default parameters. Finally, SeqMonk (https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) was used for visualization.

Detection of ASM

ASM analysis relies on the presence of heterozygous SNPs on the same read as a given CpG site to separate alleles prior to analysis. DNA methylation levels at individual CpG sites between the two alleles are then directly compared. Bisulfite-seq reads were aligned to the GRCh37/hg19 reference genome using Bowtie2. Known SNP positions in the reference genome were masked by the ambiguity base "N" prior to alignment to avoid read mapping biases. SNPsplit (https://www.bioinformatics.babraham.ac.uk/projects/SNPsplit/) was then used to determine the allelic origin of reads overlapping SNP positions using individual-specific genotype information from the gDNA-seq data. Thus, the data was split into two *in silico* allele-specific aligned genomes. The methylation status of CpGs was then individually called from the two separate alignments using Bismark and the methylation level of each CpG was compared between the two alignments. To appropriately estimate methylation differences, we required a minimum of 30 reads per allele. CpGs were defined as ASM sites if at least one individual had a percent methylation difference of $> 20\%$ between the two alleles. The individual ASM differences were weighted by their inverse variances and then summed. The sum was finally scaled by its standard error to form a Z-statistic. A p-value of less than 0.05 was considered

statistically significant. A positive percentage difference value indicates that the major allele is more methylated than the minor allele.

Evaluation of ASM Enrichment in Different Genomic Contexts

We used the right-handed Fisher's exact test to evaluate if CpGs displaying ASM were enriched in specific genomic contexts compared with the total number of CpGs in the same regions. The regions investigated for ASM enrichment were upstream regions (defined as $+5$ kb to $+2$ kb from transcription start site [TSS]), promoters (defined as $+2$ kb to -1 kb from TSS), first exon, remaining exons (excluding the first exon), introns, exon/intron boundaries, CpG islands, and CpG island shores. A p-value of less than 0.01 was considered statistically significant.

Tag-SNP Extraction from Methyl-Seq Data

Heterozygous SNPs on the same sequencing read as the CpG sites displaying ASM were extracted from the gDNA data. For CpGs with several heterozygous SNPs on the same read, the closest ones (up to three SNPs per CpG) were selected.

Association of ASM-SNPs to mRNA Expression Levels in RNA-Seq Data

Given that DNA methylation is linked to the regulation of gene expression, we assessed whether a difference in allelic mRNA expression could be identified based on the ASM-SNPs. Results from targeted RNA-seq in the same 19 individuals have previously been reported.[18] Normalized estimates of gene expression, that is, transcripts per million were calculated per gene. We then evaluated all ASM-SNPs for association to expression levels by linear regression analysis using additive models in SPSS for Windows version 20 (IBM Corporation, New York, United States). The statistical significance cutoff (two-tailed) was 0.05. As this study is based on only 19 samples, we are underpowered for interindividual analyses and therefore also report trends (defined as $p < 0.2$).

Association of ASM-SNPs to mRNA Expression in Publicly Available Expression Quantitative Trait Loci Data

We next queried all ASM-SNPs and high linkage disequilibrium (LD) proxies ($r^2 \geq 0.8$ in 1,000 genomes European panel) against publicly available expression quantitative trait loci (eQTL) data using PhenoScanner (http://www.phenoscanner.medschl.cam.ac.uk; assessed May 2019)[20] which includes results from the Genotype-Tissue Expression project v6 (http://www.gtexportal.org/home/datasets),[21,22] NHLBI GRASP v2,[23] STARNET,[24] and other smaller studies (see http://www.phenoscanner.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.

ASM-SNPs: Association with Circulating Proteins and Disease Traits

To identify ASM-SNPs associated with circulating plasma levels or activity of the respective proteins and relevant

disease and nondisease phenotypes, we queried our variants and their corresponding proxies ($r^2 \geq 0.8$) against publicly available genome-wide association study (GWAS) data using both the PhenoScanner "Proteins" and "Diseases & traits" search functions (assessed May 2019) which includes results from the NHGRI-EBI GWAS catalog (https://www.ebi.ac.uk/gwas/) and the UK Biobank (http://www.ukbiobank.ac.uk). For GWAS, summary statistics were filtered to $p < 5 \times 10^{-5}$ and then manually curated to retain only entries with relevant traits. For ASM-SNPs in *CPB2* and *HAPB2*, we supplemented this search with data from our previously published GWAS on thrombin-activatable fibrinolysis inhibitor (encoded by *CPB2*)[25] and factor-seven activating protease (FSAP, encoded by *HABP2*).[26]

Validation of ASM-SNPs in mQTL Data Sets

As a way to validate our findings, ASM-SNPs and corresponding proxies ($r^2 \geq 0.8$ in 1,000 genomes European panel) were queried against publicly available mQTL data using PhenoScanner "Epigenetics" search function which includes data from BIOS QTL browser (https://genenetwork.nl/biosqtlbrowser/), BLUEPRINT Epigenome (http://www.blueprint-epigenome.eu/), and Gaunt et al.[27] The output result list was filtered to $p < 5 \times 10^{-5}$ and manually curated to retain only mQTL entries.

Annotation and Functional Prediction of Variants

Database of Single Nucleotide Polymorphisms (dbSNP; build 150, available from http://www.ncbi.nlm.nih.gov/SNP/) was used to annotate SNPs. HaploReg v4.1[28] was used to determine the chromatin state in liver tissue per ASM-SNP haplotype block. The default 15-state core ChromHMM model and H3K4me1/H3K4me3 peaks from Roadmap Epigenomics[29] were used to identify enhancers and promoters. Results were manually curated to retain only variants with predicted enhancer or promoter chromatin states in liver tissue.

Results

DNA Methylation Pattern and ASM of Hemostatic Genes in Human Liver

We measured the DNA methylation status of each individual CpG in all 35 hemostatic genes including at least 5 kb upstream and 0.5 kb downstream of each gene. In total, 8,707 unique CpG sites were captured in our design (1.2 Mb in total). The density of CpG sites in each gene varied between 2 CpGs (*F13B*) and 30 CpGs (*F7*) per kbp, with a median density of 6 CpGs per kbp. Schematic illustrations of DNA methylation patterns for *VKORC1* (encoding vitamin K reductase complex subunit 1), *F2* (thrombin), and *F10* (factor X) are displayed in ►**Fig. 1** and for the remaining genes in ►**Supplementary Fig. S2** (available in the online version).

Using the ASM approach, we identified 151 CpGs displaying an allelic methylation imbalance of 20% or more in at least one sample, and 116 of these were significant ($p < 0.05$). These CpGs were found in 24 of the 35 genes analyzed and were associated with 112 SNPs (►**Table 1**). The number of CpGs displaying ASM per gene varied between one (e.g., in *VKORC1*)

to 15 (e.g., in *SERPINF2*). A schematic view of ASM positions per gene is displayed in ►**Fig. 1** and in ►**Supplementary Fig. S2** (available in the online version). The exact positions of the ASM-CpGs, the degree of ASM, dbSNP ID of the ASM-SNPs, number of heterozygous samples, and associations of ASM-SNPs with mRNA expression, levels/activity of circulating proteins, and other GWAS traits with relevance for hemostasis are reported in ►**Tables 2–4**. The ASM association results are presented in these three tables based on whether or not the ASM-SNPs have previously been identified as eQTLs. Allelic methylation for all CpGs (including nonsignificant ASM-CpGs) and further details on associations are shown in ►**Supplementary Table S3** (available in the online version). Forest plots of methylation differences between alleles in all 151 CpGs per individual are displayed in ►**Supplementary Fig. S3** (available in the online version). A UCSC browser track for significant ASM positions is available to the research community at: https://genome.ucsc.edu/s/marcela.davila%40gu.se/ASM_CJern.

CpGs Displaying ASM are Enriched in the First Exon

To evaluate if ASM is overrepresented in specific genomic contexts, we performed an enrichment test comparing ASM-CpGs with all CpGs in different regions. The only region with a significant over-representation of ASM was "first exon" ($p = 0.0007$, Fisher's exact test).

### Eight ASM-SNPs Associated with mRNA Expression in RNA-Seq Data

To determine whether DNA methylation regulated by genetic variants in *cis* could provide the mechanistic link between genetic variants associated with gene expression, we used mRNA-seq data (derived from the same individuals and the same liver tissue samples as used in ASM analysis) to assay potential differences in expression based on genotype of the ASM-SNP. A total of eight ASM-SNPs were significantly associated with allelic differences in expression ($p < 0.05$): four in the gene encoding protein Z (PROZ), one in the gene encoding VKORC1, one in the gene encoding thrombin, and two in the gene encoding kininogen 1 (KNG1) (►**Fig. 2**; ►**Tables 2** and **3**). A trend was observed for a further 22 ASM-SNPs (9 at $p < 0.1$ and 13 ASM-SNPs at $p < 0.2$; see online ►**Supplementary Fig. S4**, available in the online version).

Seventy-Six ASM-SNPs Associated with mRNA Expression in Publicly Available eQTL Data

We next queried the ASM-SNPs in publicly available liver tissue eQTL data using PhenoScanner. A total of 22 ASM-SNPs (or high LD proxies) were identified as eQTLs in liver tissue for their respective gene (►**Table 2**). These included four SNPs in *PROZ* and one SNP in *VKORC1* for which we found a significant association with mRNA expression using our RNA-seq data (►**Fig. 2A–D** and **E**, respectively) and five others with a trend for association. The same alleles associated with higher gene expression in liver tissue eQTL studies were expressed at higher levels in our RNA-seq data (data not shown).

While many *cis*-acting genetic variants are specific for a given tissue, others may be common to several tissues.[22]
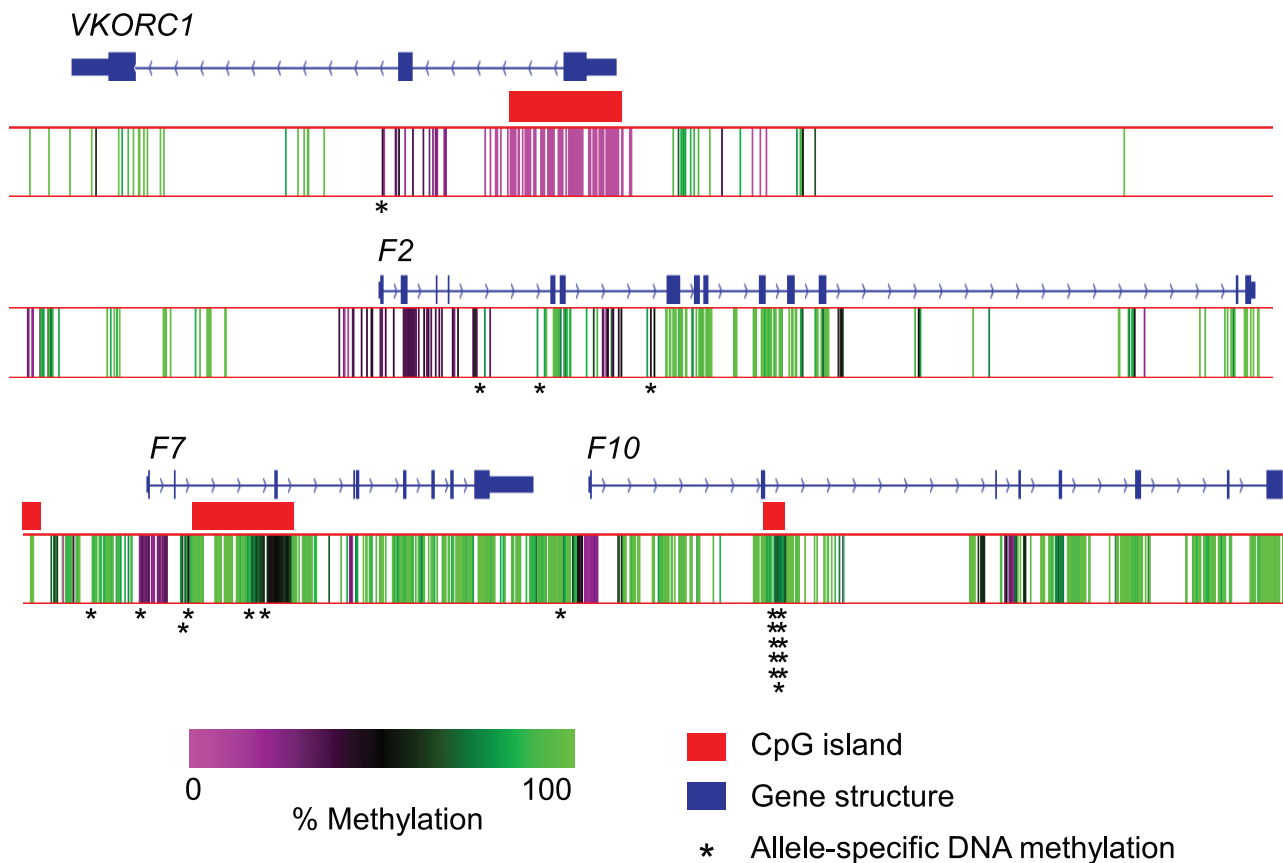
Fig. 1 Deoxyribonucleic acid (DNA) methylation patterns of selected hemostatic genes in human liver tissue. For each gene, gene structure (blue) and cytosine–guanine dinucleotide (CpG) islands (red) are displayed on top. At least 5 kb upstream and 0.5 kb downstream of each gene are included. The methylation status for all CpGs is displayed, where purple represents a low degree of methylation and green a high degree of methylation. Methylation values are based on the average of the 19 samples. CpGs displaying allele-specific methylation (ASM) in one or more samples are denoted by asterisks. (A) VKORC1; (B) F2; and (C) F7 and F10. All other genes can be found in ▶Supplementary Fig. S2 (available in the online version).

Further, sample size greatly affects eQTL mapping and thus more readily available tissues tend to have more reported eQTLs.[22] In line with this, an additional 54 ASM-SNPs (or proxies) were eQTLs for their respective genes in other tissues in a total of 17 genes (▶Table 3). This included four SNPs in *PROZ*, one in *F2*, and two in *KNG1*, significantly associated with expression using our normalized RNA-seq data (▶Fig. 2F–H), and 13 other SNPs showing a trend for association in our data (▶Supplementary Fig. S4, available in the online version). A complete list of all eQTL associations from all tissues can be found in ▶Supplementary Table S4 (available in the online version).

For 36 ASM-SNPs (corresponding to 16 genes), no genotype-expression associations have previously been reported (▶Table 4). Of these, three ASM-SNPs exhibited a trend for association with expression using our normalized RNA-seq data for *PROZ* and the genes encoding factor V (F5) and α-2-plasmin inhibitor (SERPINF2).

Thirty-Three ASM-SNPs Associate with Circulating Proteins; 36 with Other Relevant GWAS Traits

To determine whether allele-specific DNA methylation also could affect protein abundance, we assessed ASM-SNPs for association with circulating levels or activity of plasma proteins using publicly available protein quantitative trait loci (pQTL) and GWAS data sets. A total of 33 ASM-SNPs (or proxies) were previously associated with the respective protein concentration or activity (▶Tables 2–4, details in ▶Supplementary Table S3, available in the online version), and these included 5 ASM-SNPs with no previous reported eQTLs that were associated with plasma fibrinogen, KNG1, SERPINF2, and FSAP levels (▶Table 4). A complete list of pQTL associations can be found in ▶Supplementary Table S5 (available in the online version).

We next assessed association of ASM-SNPs with relevant disease and nondisease phenotypes in PhenoScanner. One ASM-SNP in *VKORC1* was associated with warfarin dosage levels (rs8050894).[30] A total of four ASM-SNPs were associated with complex diseases (two SNPs in *F5* with venous thromboembolism [rs3766110, rs3766111; http://www.ukbiobank.ac.uk], one SNP in *TFPI* with coronary artery disease [rs8176546],[31] and one in *F2* with stroke [rs2070851][32]). A total of 33 ASM-SNPs were associated with other related phenotypes. A complete list of GWAS associations can be found in ▶Supplementary Table S5 (available in the online version).

Table 1 Summary of ASM CpG-SNP results from integrated targeted sequencing of 35 hemostatic genes in human liver, separated based on whether prior eQTLs for the respective SNP have been documented

	No. CpG-SNP pairs	No. ASM-CpGs	No. ASM-SNPs	Genes with ASM No.	Genes with ASM Gene names	ASM-SNPs with pQTL No.	ASM-SNPs with pQTL Gene names	ASM-SNPs in promoter or enhancer[a]
eQTL in liver (→**Table 2**)	31	30	22	9	C4BPB, CPB2, F7, F10, PROC, PROZ, SERPINA1, TFPI, VKORC1	11	CPB2, F7, F10, PROC	21
eQTL other tissues (→**Table 3**)	71	61	54	17	A2M, CPB2, F2, F5, F10, F13B, FGB, KNG1, PLG, PROC, SERPINA1, SERPINA5, SERPINA10, SERPIND1, SERPINF2, SERPING1, TFPI	17	F5, F10, F13B, FGB, CPB2, KNG1, PROC, SERPINA10, SERPINF2, SERPING1	54
No eQTL reported (→**Table 4**)	42	36	36	16	F2, F5, F7, F10, F13B, FGB, HABP2, HRG, KNG1, PLG, PROC, PROS1, PROZ, SERPINA5, SERPIND1, SERPINF2	5	FGB, HABP2, KNG1, SERPINF2	27
Total	144	116	112	24 genes		33	12 genes/proteins	105

Abbreviations: ASM, allele-specific DNA methylation; CpG, cytosine-guanine site; DNA, deoxyribonucleic acid; eQTL, expression quantitative trait loci; No. ASM-CpGs, number of CpG positions displaying ASM; No. ASM-SNPs, number of SNPs associated with CpGs displaying ASM; No. CpG-SNP pairs, number of CpG-SNP pairs; pQTL, protein quantitative trait loci; SNP, single-nucleotide polymorphism.
[a]Predicted promoter (including intergenic promoters) or enhancer based on data from HaploReg.

Overlap between ASM-SNPs and Previously Identified mQTL SNPs

We queried our ASM-CpGs against the publicly available mQTL microarray-based data using PhenoScanner. Only six of the CpGs displaying ASM in our study design are located on commercially available microarrays. However, all six of these were identified as mQTLs with the same SNP-CpG association as in our study, although in other tissue types than liver. Therefore, all remaining ASM-CpGs identified here are completely novel genotype-associated methylation sites, and all associations reported in this study are novel in liver tissue.

Discussion

In this study, we analyzed DNA methylation profiles of 35 hemostatic genes that are predominantly expressed in liver. To the best of our knowledge, this is the first study presenting DNA methylation patterns of hemostatic genes on an individual CpG level in human liver. We identified ASM in 24 of the 35 targeted genes. In addition, we demonstrated associations between ASM-SNPs and mRNA expression as well as association with circulating hemostatic proteins, relevant complex traits (e.g., coronary artery disease, stroke, venous thromboembolism), and drug dosing (e.g., warfarin). Our findings highlight the importance of genetic–epigenetic interactions and suggest that methylation regulated by genetic variants could provide the mechanism by which SNPs exert an effect on circulating hemostatic proteins, prothrombotic diseases, and drug response.

As discussed in the Introduction, previous studies that link variations in genotypes to DNA methylation levels at specific CpG sites across individuals (mQTL) require large sample sizes and have most often been performed in easy to access tissues such as blood. Here, we used ASM analysis, which has an advantage over mQTL analyses as it uses a within-sample control (the other allele) to query effects of SNPs on DNA methylation. In doing so, the ASM approach is more robust to environmental and experimental effects and thus more sensitive than standard mQTL analysis.[33] We were therefore able to use human liver tissue, the main biological source of the hemostatic proteins studied here, to map putative *cis*-acting polymorphisms despite having a small number of samples.[17] ASM-SNPs and corresponding proxies were queried against publicly available mQTL data as a way to validate our findings. However, as nearly all published human mQTL studies have used commercially available microarrays that interrogate only a small portion of all CpGs, only eight CpGs in our study design are represented in mQTL data. Six of these displayed ASM in our data, and all of these were identified as mQTLs with the same SNP-CpG association as in our study, although in other tissue types than liver. Ideally, we would compare our ASM results with another similar bisulfite-seq data set in liver, but such data is not available at this time as far as we are aware.

The hemostatic genes analyzed in this study all had a low degree of methylation around promoters, as expected for actively transcribed genes (►**Fig. 1** and

Table 2 CpGs displaying significant ASM ($p < 0.05$) and associated SNPs (ASM-SNPs) previously reported as liver eQTLs in public databases, and associations of the ASM-SNPs to mRNA expression in liver tissue, to circulating protein levels of the respective genes, and to other GWAS traits relevant for hemostasis

	ASM			SNP		Associations		
Gene	CpG location	No. het.	% ASM diff.	dbSNP	Distance from CpG (bp)	mRNA	pQTL	GWAS
C4BPB	chr1:207269856	1	−21	rs8942	63	Yes[b]		
CPB2	chr13:46644103	2	9	rs9534307	27	Yes[b]	Yes	
CPB2	chr13:46649378	1	−27	rs9567613	20	Yes[b]	Yes	
CPB2	chr13:46669605	1	−21	rs7991899	−55	Yes[b]	Yes	
CPB2	chr13:46683614	2	−10	rs7338061	19	Yes[b]	Yes	
F10	chr13:113784439[a]	1	23	rs693335	4		Yes	Yes
F10	chr13:113784452	1	20	rs693335	−9		Yes	Yes
F10	chr13:113784460	1	25	rs693335	−17		Yes	Yes
F10	chr13:113784466	1	26	rs693335	−23		Yes	Yes
F10	chr13:113784481	1	22	rs693335	−38		Yes	Yes
F10	chr13:113784497	1	23	rs693335	−54		Yes	Yes
F10	chr13:113784499	1	25	rs693335	−56		Yes	Yes
F7	chr13:113757759	2	−12	rs3093229	67		Yes	Yes
F7	chr13:113759695	1	−25	rs510317	59		Yes	Yes
F7	chr13:113761451	2	24	rs28663357	62	Yes[b]		Yes
F7	chr13:113761585[a]	2	15	rs142493427	5			Yes
PROC	chr2:128174852	1	−28	rs2069901	−9		Yes	
PROC	chr2:128174852	1	−28	rs2069902	11		Yes	
PROC	chr2:128176028	9	−6	rs1799810	12		Yes	
PROC	chr2:128177360	8	11	rs1158867	17		Yes	
PROZ	chr13:113817559	1	45	rs2480948	−22	Yes	NA	Yes
PROZ	chr13:113818668	2	−24	rs3024731	40	Yes	NA	Yes
PROZ	chr13:113819002	5	26	rs494860	7	Yes	NA	Yes
PROZ	chr13:113819008	5	27	rs494860	1	Yes	NA	Yes
PROZ	chr13:113819050	2	29	rs494860	−41	Yes	NA	Yes
PROZ	chr13:113819074[a]	2	18	rs494860	−65	Yes	NA	Yes
PROZ	chr13:113824222	1	32	rs1755690	44	Yes	NA	Yes
SERPINA1	chr14:94844610	1	−24	rs2073333	−48		NA	
SERPINA1	chr14:94844885	1	31	rs1303	−42		NA	
TFPI	chr2:188340376	4	−11	rs8176546	20		NA	Yes
VKORC1	chr16:31104505	3	−47	rs8050894	4	Yes	NA	Yes

Abbreviations: ASM, allele-specific DNA methylation; bp, base pairs; CpG, cytosine-guanine site; dbSNP, Database of Single Nucleotide Polymorphisms; DNA, deoxyribonucleic acid; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; mRNA, messenger ribonucleic acid; pQTL, protein quantitative trait loci; SNP, single-nucleotide polymorphism.

Note: No. het., number of heterozygous samples; % ASM diff., mean percent methylation difference between major and minor alleles of all heterozygous samples. A positive percent difference indicates higher methylation level on the major allele.

[a]CpG included on commercial DNA methylation microarrays.

[b]Nonsignificant ($p > 0.05$) trend for association with mRNA expression in in-house mRNA-seq data.

►**Supplementary Fig. S2**, available in the online version). The methylation level in gene bodies was generally much higher, although this varied more than the promoter methylation (►**Fig. 1** and ►**Supplementary Fig. S2**, available in the online version). Previous mQTL studies in a variety of tissues have demonstrated that *cis*-acting methylation-associated SNPs are enriched in regulatory chromatin domains and transcription factor binding sites, and colocalize with eQTLs.[11,34–37] Further, CpG sites that are associated with expression levels have previously been shown to be enriched in enhancers,

Table 3 CpGs displaying significant ASM ($p < 0.05$) and associated SNPs (ASM-SNPs) previously reported as eQTLs in tissue types other than liver, and association of these ASM-SNPs to mRNA expression in liver tissue (RNA-seq data), to circulating protein levels of the respective genes (pQTL), and to other GWAS traits relevant for hemostasis

Gene	ASM			SNP		Associations		
	CpG location	No. het.	% ASM diff.	dbSNP	Distance from CpG (bp)	mRNA	pQTL	GWAS
A2M	chr12:9230027	7	−18	rs2377682	11		NA	
A2M	chr12:9259915	6	−10	rs226389	−8	Yes[b]	NA	
A2M	chr12:9279617	7	−6	rs226372	−19	Yes[b]	NA	
CPB2	chr13:46627428	8	−28	rs1087	11	Yes[b]	Yes	
CPB2	chr13:46627428	8	−28	rs940	52		Yes	
CPB2	chr13:46627473	4	−23	rs940	7		Yes	
CPB2	chr13:46627504	3	−12	rs940	−24		Yes	
CPB2	chr13:46649378	1	−27	rs1022952	−30		Yes	
F10	chr13:113784564	1	36	rs776906	24		Yes	Yes
F10	chr13:113784692	1	21	rs2480946	28		Yes	Yes
F10	chr13:113784702	1	−24	rs2480946	18		Yes	Yes
F10	chr13:113776916[a]	4	−11	rs563964	31	Yes[b]		Yes
F10	chr13:113776916[a]	4	−11	rs3211717	33	Yes[b]	Yes	
F13B	chr1:197026831	10	8	rs17549312	15		Yes	
F13B	chr1:197026871	6	12	rs17549312	−25		Yes	
F2	chr11:46744514	2	17	rs2070851	14		NA	Yes
F2	chr11:46747073	1	−50	rs3136460	−8	Yes	NA	Yes
F5	chr1:169493712	1	25	rs9332641	1		Yes	Yes
F5	chr1:169493712	1	25	rs9332640	14		Yes	
F5	chr1:169515148	1	−27	rs3766110	35			Yes
F5	chr1:169515148	1	−27	rs3766111	56			Yes
F5	chr1:169533986	1	21	rs724507	42			
F5	chr1:169555972	1	−25	rs2269648	78			
FGB	chr4:155486657	3	12	rs2227403	−11		Yes	
KNG1	chr3:186435526	1	−22	rs13072823	35	Yes		Yes
KNG1	chr3:186440466	2	16	rs1648714	17		Yes	Yes
KNG1	chr3:186449900	6	−6	rs1622922	−16	Yes[b]		Yes
KNG1	chr3:186449900	6	−6	rs1648699	−10	Yes[b]		Yes
KNG1	chr3:186449900	6	−6	rs1648700	16	Yes[b]		Yes
KNG1	chr3:186452968	2	−19	rs822363	23			Yes
KNG1	chr3:186434963	2	−16	rs5029973	−76	Yes		Yes
PLG	chr6:161135056	2	11	rs3757019	55	Yes[b]		
PLG	chr6:161137753	2	16	rs14224	26	Yes[b]		
PROC	chr2:128182990	8	29	rs2069924	4		Yes	
PROC	chr2:128147838	7	8	rs13029237	16		Yes	
PROC	chr2:128148742	6	−22	rs35708549	22			
PROC	chr2:128148776	8	−27	rs35708549	−12			
SERPINA1	chr14:94843725	1	23	rs11628917	−6		NA	Yes
SERPINA1	chr14:94852549	1	25	rs1980617	−15		NA	Yes
SERPINA1	chr14:94852749	2	−16	rs75285293	47		NA	

Table 3 (Continued)

Gene	ASM CpG location	No. het.	% ASM diff.	SNP dbSNP	Distance from CpG (bp)	mRNA	pQTL	GWAS
SERPINA1	chr14:94858368	2	−12	rs1610317	−12		NA	Yes
SERPINA10	chr14:94755152	2	−22	rs12434093	32		Yes	
SERPINA10	chr14:94755217	1	−21	rs12434093	−33		Yes	
SERPINA10	chr14:94755230	1	−32	rs12434093	−46		Yes	
SERPINA10	chr14:94760975	5	−9	rs3827897	−10		Yes	
SERPINA5	chr14:95050296	2	−22	rs12880862	12			
SERPINA5	chr14:95051504	1	−40	rs12886695	−64			
SERPINA5	chr14:95047617	2	−12	rs11847476	−12	Yes[b]		
SERPINA5	chr14:95047617	2	−12	rs11847477	−9			
SERPIND1	chr22:21129590	1	25	rs165680	−31	Yes[b]	NA	
SERPINF2	chr17:1648204	1	−32	rs2070862	90			
SERPINF2	chr17:1650172	1	−40	rs117378188	31	Yes[b]		
SERPINF2	chr17:1658948	1	−23	rs4566204	20			
SERPINF2	chr17:1658955	1	−26	rs4566204	13			
SERPINF2	chr17:1636455	1	−29	rs11657222	22	Yes[b]		
SERPINF2	chr17:1640652[a]	1	46	rs62090051	−12			
SERPINF2	chr17:1640661	1	36	rs62090051	−21			
SERPINF2	chr17:1640674	1	44	rs62090051	−34			
SERPINF2	chr17:1640674	1	44	rs1045794	35	Yes[b]		
SERPINF2	chr17:1640710	1	−30	rs62090051	−70			
SERPINF2	chr17:1640710	1	−30	rs1045794	−1	Yes[b]		
SERPINF2	chr17:1640788	11	−7	rs8077638	5		Yes	
SERPINF2	chr17:1640794	1	−31	rs1045794	−85	Yes[b]		
SERPINF2	chr17:1640794	1	−31	rs8077638	−1		Yes	
SERPINF2	chr17:1640812	10	−10	rs8077638	−19		Yes	
SERPINF2	chr17:1640830	4	−18	rs8077638	−37		Yes	
SERPINF2	chr17:1640868	1	−21	rs8077638	−75		Yes	
SERPINF2	chr17:1645507	5	−14	rs62090057	19			
SERPING1	chr11:57374280	1	−22	rs11603020	52	Yes[b]	Yes	
TFPI	chr2:188340376	4	−11	rs8176547	−27		NA	
TFPI	chr2:188353789	1	36	rs8176500	−35		NA	

Abbreviations: ASM, allele-specific DNA methylation; bp, base pairs; CpG, cytosine-guanine site; dbSNP, Database of Single Nucleotide Poly-morphisms; DNA, deoxyribonucleic acid; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; mRNA, messenger ribonucleic acid; pQTL, protein quantitative trait loci; SNP, single-nucleotide polymorphism.
Note: No het, number of heterozygous samples; % ASM diff., mean percent methylation difference between major and minor alleles of all heterozygous samples. A positive percent difference indicates higher methylation level on the major allele.
[a]CpG included on commercial DNA methylation microarrays.
[b]Nonsignificant ($p > 0.05$) trend for association with mRNA expression in in-house RNA-seq data.

gene bodies, CpG island shores, and in the first exon, but not in promoter regions.[8,38] In line with this, we found that the majority of the ASM-CpGs/SNPs were located within gene bodies (i.e., 102 out of 144) and identified an enrichment of ASM in the first exon. Additionally, nearly all (94%) of the ASM-SNPs had chromatin states indicative of promoter and/or enhancer elements in liver (►**Supplementary Table S6**, available in the online version). Taken together, we present the first detailed map of the DNA methylation landscape in hemostatic genes in human liver tissue, and our study underscores the importance of more mechanistic studies on the role of methylation in gene bodies.

Table 4 CpGs displaying significant ASM ($p < 0.05$) and associated SNPs (ASM-SNPs) not reported as eQTLs in public databases, and associations of ASM-SNPs to mRNA expression in liver tissue (RNA-seq data), to circulating protein levels of the respective genes (pQTL), and to other GWAS traits relevant for hemostasis
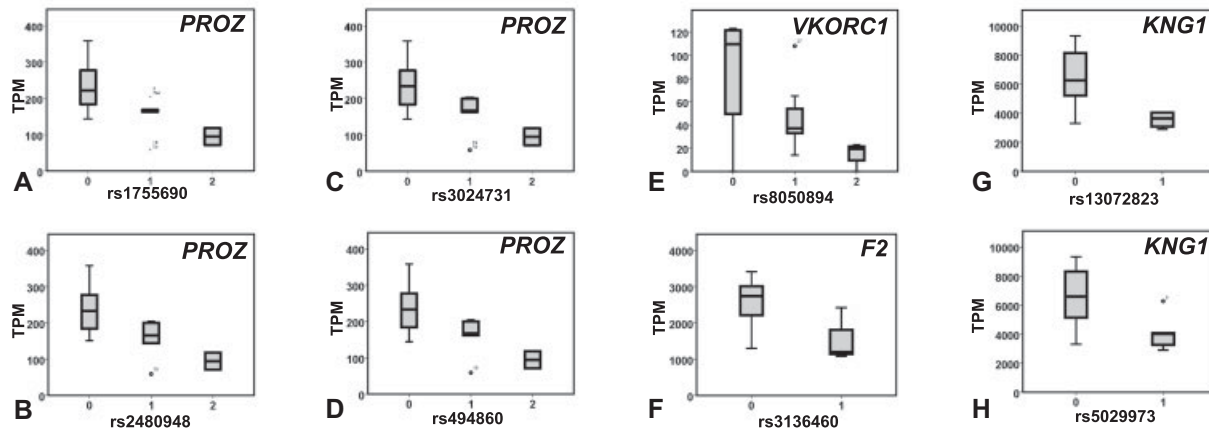
	ASM			SNP		Associations		
Gene	CpG location	No. het.	% ASM diff.	dbSNP	Distance from CpG (bp)	mRNA	pQTL	GWAS
F10	chr13:113784664	1	31	rs3211742	−20			
F13B	chr1:197026871	6	12	rs1759008	24			Yes
F2	chr11:46743111	4	17	rs3136439	24		NA	
F5	chr1:169515447	1	42	rs749045929	−29			
F5	chr1:169515447	1	42	rs3835453	−25			
F5	chr1:169515447	1	42	rs796545997	−27	Yes[b]		Yes
F5	chr1:169515447	1	42	rs112177003	−1			
F5	chr1:169529841	1	24	rs6022	−15			
F7	chr13:113759695[a]	1	−25	rs510335	60			
F7	chr13:113761585[a]	2	15	rs2515646	17			
FGB	chr4:155480425	1	21	rs7673587	9		Yes	
HABP2	chr10:115310488	5	−9	rs2000277	26			
HABP2	chr10:115310548	6	12	rs3758591	4		Yes	
HABP2	chr10:115310548	6	12	rs3758590	15			
HABP2	chr10:115316814	1	32	rs4918842	−2			
HABP2	chr10:115327006	1	23	rs11575665	34			
HABP2	chr10:115327006	1	23	rs11575666	42			
HABP2	chr10:115330374	7	56	rs7079148	12			
HABP2	chr10:115332849	10	33	rs7101144	1			
HABP2	chr10:115334080	2	13	rs3740530	44		Yes	
HABP2	chr10:115336796	1	24	rs2302374	−38			
HRG	chr3:186383214	1	−21	rs9814412	−34			Yes
KNG1	chr3:186442126	1	−20	rs1648712	83			Yes
KNG1	chr3:186445048	2	16	rs2304456	4		Yes	Yes
KNG1	chr3:186445221	2	30	rs3841557	−35			
KNG1	chr3:186445221	2	30	rs5030025	16			Yes
PLG	chr6:161126010	1	−22	rs1819138	64			
PLG	chr6:161126030	1	−35	rs1819138	44			
PLG	chr6:161168694	1	−27	rs10625439	−41			
PROC	chr2:128148742	6	−22	rs79819102	−1			
PROC	chr2:128148776	8	−27	rs79819102	−35			
PROS1	chr3:93605725	1	−45	rs8178648	14		NA	
PROZ	chr13:113817557	1	49	rs45538837	−2		NA	Yes
PROZ	chr13:113817559	1	45	rs45538837	−4		NA	Yes
PROZ	chr13:113818958	2	16	rs3024733	44		NA	
PROZ	chr13:113819002	5	26	rs3024733	0		NA	
PROZ	chr13:113819008	5	27	rs3024733	−6		NA	
PROZ	chr13:113819093	1	23	rs3024782	65	Yes[b]	NA	
PROZ	chr13:113819101	1	25	rs3024782	57	Yes[b]	NA	
SERPINA5	chr14:95048706	1	−27	rs112795992	−1			
SERPIND1	chr22:21132447	2	−32	rs75917143	−9		NA	
SERPINF2	chr17:1650172	1	−40	rs6502935	−4	Yes[b]	Yes	

Abbreviations: ASM, allele-specific DNA methylation; bp, base pairs; CpG, cytosine-guanine site; dbSNP, Database of Single Nucleotide Polymorphisms; DNA, deoxyribonucleic acid; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; mRNA, messenger ribonucleic acid; pQTL, protein quantitative trait loci; SNP, single-nucleotide polymorphism.
Note: No. het., number of heterozygous samples; % ASM diff., mean percent methylation difference between major and minor alleles of all heterozygous samples. A positive percent difference indicates higher methylation level on the major allele.
[a]CpG included on commercial DNA methylation microarrays.
[b]Nonsignificant ($p > 0.05$) trend for association with mRNA expression in in-house RNA-seq data.

Fig. 2 Box plots of genotype versus messenger ribonucleic acid (mRNA) expression level. Allele-specific methylation-single-nucleotide polymorphisms (ASM-SNPs) significantly associated to expression of the respective gene based on normalized RNA-seq data from 19 human liver samples ($p < 0.05$). Subjects homozygous for the major allele had higher gene expression of PROZ (**A–D**); VKORC1 (**E**); F2 (**F**); and KNG1 (**G, H**). TPM, transcripts per million. 0 = homozygous major allele, 1 = heterozygous, 2 = homozygous minor allele. Plots for ASM-SNPs showing a trend for association to expression ($p < 0.2$) can be found in ►**Supplementary Fig. S2** (available in the online version).

We found that 68% of our ASM-SNPs had previously been associated with expression of their respective genes in publicly available eQTL databases. However, only a minority (22 of 76) were liver eQTLs. Thus, the ASM-SNPs were also tested for association to expression using RNA-seq data from our liver samples. We report both significant associations and trends. We are aware that such an analysis should ideally be adjusted for multiple testing. However, as we were underpowered for inter-individual analysis, we believe this is an appropriate first step as we used the tissue of relevance.[39] In addition to coming from the same individuals, DNA methylation and RNA analyses were analyzed from the same tissue samples (i.e., DNA and RNA were extracted simultaneously, with the same exact cell composition) in our study which can facilitate the investigation of the functional relationship between DNA methylation and gene expression.[40] In support of our RNA-seq results, all significant associations were previously reported eQTLs (five in liver and three in other tissues). Of the 22 ASM-SNPs with a trend for association, 19 were previously reported eQTLs (6 in liver and 13 in other tissues). Therefore, despite using so few samples, our study supports many published eQTLs for hemostatic genes, even those initially identified in other tissues, and provides validity to our study design.

eQTLs tend to be found in promoter regions, whereas mQTLs and cis-ASM positions tend to occur in other regulatory regions such as enhancers and insulators.[33] Previous studies have found only a small overlap between variants tagging cis-acting DNA methylation and expression (i.e., mQTLs and eQTLs).[11,15,27,41–43] We recently performed a genotype-expression study by analyzing allele-specific expression (ASE) of the same 35 hemostatic genes in the same sample used in the present study.[18] We compared all ASM-SNPs and ASE-SNPs (including corresponding proxies) identified in both studies and found only 10 SNPs in common ($< 1\%$). As ASE analyses are limited to exonic SNPs (i.e., SNPs

need to be present on RNA), whereas ASM is based on CpG sites that are located primarily in upstream regulatory regions and introns, these two analyses represent complementary approaches to identify novel cis-regulatory variants. Taken together, the results from our previous and present study highlight the complimentary nature of conducting both types of mapping.

Given the range of processes that affect protein abundance, a SNP influencing DNA methylation and mRNA levels may not necessarily influence the protein level. To address whether DNA methylation regulated by SNPs in cis could provide the mechanistic link between SNPs previously associated with circulating hemostatic proteins, we intersected the ASM-SNPs (or high LD proxies) with public pQTL and GWAS data. As far as we are aware, pQTL or GWAS protein data (for SNPs in the respective gene) is only available for 16 of the 24 genes displaying ASM in our study. For 12 of these genes (75%), the SNPs associated with protein levels in the public data were the same as the ASM-SNPs (associated with CpG methylation) identified here. Although this analysis suggests a potential role for many of the ASM-SNPs in regulation of plasma proteins, it should be emphasized that further replication of our results and other functional validation studies are warranted.

In the era of personalized therapy, several clinical trials are investigating the value of genotype-guided dosing of the most commonly used oral anticoagulants, for example, warfarin, and nonvitamin K antagonist oral anticoagulants (NOACs).[44–46] Studies are also emerging that demonstrate that DNA methylation changes contribute to the variability of response to warfarin,[47] aspirin, and clopidogrel.[48] Interestingly and in line with this, we identified ASM CpG-SNP pairs in VKORC1 which is a target for warfarin, and in F2 and F10 which both are targets for NOACs. We also identified ASM CpG-SNP pairs in 11 other hemostatic genes that are current targets of approved antithrombotic, antifibrinolytic, and other related drugs.[49] The ASM-SNP in VKORC1 is an eQTL

SNP in liver for VKORC1 expression, was associated with mRNA expression in our study ($p = 0.009$), and is the top hit in a GWAS for "warfarin maintenance dose."[30] The ASM-CpG associated with this SNP displayed one of the highest degrees of ASM in our study, with a percent methylation difference of 47% in favor of the minor allele. A SNP in almost perfect LD with the ASM-SNP has already been used in pharmacogenetics-based clinical trials for warfarin with promising results in European populations.[44] Thus, methylation influenced by genetic variation in *cis* may not only provide a mechanistic link between the noncoding genetic changes and phenotypic variation observed in circulating hemostatic proteins and prothrombotic diseases but also in response to drugs.

Apart from the methodological strengths of the ASM approach described above, we used N-masked individualized genomes based on gDNA-seq data to prevent alignment bias which is otherwise a common source of error in ASM analysis. Additionally, the custom target sequence design used here provides much deeper coverage across the genes of interest compared with whole genome bisulfite-seq. Despite these strengths, there are some limitations which deserve mention. (1) The targeted design we chose was limited to approximately 5 kb upstream and 0.5 kb downstream of each gene, which means we could have missed important SNPs or CpGs outside of these regions. However, it is of note that our design still provides a great enrichment compared with commonly used array-based analyses. (2) Although the ASM approach is powerful, our study is based on just 19 samples and increasing the number of samples may reveal many more CpG-SNP associations. (3) We used conventional (short-read) Illumina sequencing. To assess for ASM, one must obtain both allelic and methylation signals on the same sequencing read. Therefore, using a platform that produces longer reads (e.g., single-molecule real-time bisulfite sequencing)[50] may reveal further CpG-SNP associations. (4) Liver tissue consists of multiple cell types that may have different methylation profiles, and we recognize that we cannot determine from which cell type the identified allelic imbalances originates using this study design. (5) Additionally, it should be emphasized that the results were generated in samples from subject with European ancestry, and our findings many not be generalizable to populations of different ethnicity.

In conclusion, we identified 112 genetic variants associated with DNA methylation in hemostatic genes in human liver tissue. Of these ASM-SNPs, 78% were associated with expression of their respective genes, or with their respective protein levels or activity, or with other relevant hemostatic GWAS traits, suggesting that methylation regulated by genetic variants may provide a mechanistic link between noncoding SNPs and variation observed in circulating hemostatic proteins, prothrombotic diseases, and drug response. Our study highlights the importance of interaction analysis between genetic and epigenetic variation in relevant tissues, and that this approach can contribute to new insights into the biological processes affecting thrombosis and hemostasis.

What is known about this topic?

- Characterizing the complex relationship between genetic, epigenetic (e.g., DNA methylation), and transcriptomic variation has the potential to increase understanding about the mechanisms regulating hemostasis and inform drug development.
- Several hemostatic factors are synthesized in the liver. Despite DNA methylation being the most studied epigenetic trait, most studies are based on arrays that cover < 4% of CpG sites. Thus, no high-resolution information on DNA methylation is currently available for hemostatic genes in liver.
- Single-nucleotide polymorphisms (SNPs) can exert an influence on DNA methylation in *cis* which can affect gene expression. This can be analyzed through allele-specific methylation (ASM) experiments.

What does this paper add?

- We used a targeted bisulfite-sequencing approach to allow for the unbiased detection of all CpGs across 35 hemostatic genes and present the first high-resolution DNA methylation map for these genes in human liver.
- We identified 112 candidate ASM-SNPs in 24 hemostatic genes, all representing novel genotype-methylation associations in liver tissue. Of these SNPs, 76 were previously associated with gene expression and 61 were previously associated with circulating hemostatic proteins, relevant complex traits (e.g., coronary artery disease, stroke, venous thromboembolism), and drug dosing (e.g., warfarin).
- This suggests that methylation regulated by SNPs in *cis* may provide a mechanistic link between noncoding SNPs and variation observed in hemostatic gene expression, circulating hemostatic proteins, prothrombotic diseases, and drug response.

References

1 Kerkel K, Spadola A, Yuan E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet 2008;40(07):904–908

2 Zhang Y, Rohde C, Reinhardt R, Voelcker-Rehage C, Jeltsch A. Non-imprinted allele-specific DNA methylation on human autosomes. Genome Biol 2009;10(12):R138

3 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

4 Paliwal A, Temkin AM, Kerkel K, et al. Comparative anatomy of chromosomal domains with imprinted and non-imprinted allele-specific DNA methylation. PLoS Genet 2013;9(08):e1003622

5 Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem 2005;74:481–514

6 Robertson KD. DNA methylation and human disease. Nat Rev Genet 2005;6(08):597–610

7 Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009;462(7271):315–322

8 Brenet F, Moh M, Funk P, et al. DNA methylation of the first exon is tightly linked to transcriptional silencing. PLoS One 2011;6(01):e14524

9 Aran D, Toperoff G, Rosenberg M, Hellman A. Replication timing-related and gene body-specific methylation of active human genes. Hum Mol Genet 2011;20(04):670–680

10 Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. On the presence and role of human gene-body DNA methylation. Oncotarget 2012;3(04):462–474

11 Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol 2014;15(02):R37

12 Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell 2014;26(04):577–590

13 Kurdyukov S, Bullock M. DNA methylation analysis: choosing the right method. Biology (Basel) 2016;5(01):3

14 Kint S, De Spiegelaere W, De Kesel J, Vandekerckhove L, Van Criekinge W. Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. PLoS One 2018;13(06):e0199091

15 Grundberg E, Meduri E, Sandling JK, et al; Multiple Tissue Human Expression Resource Consortium. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet 2013;93(05):876–890

16 Kaplow IM, MacIsaac JL, Mah SM, McEwen LM, Kobor MS, Fraser HB. A pooling-based approach to mapping genetic variants associated with DNA methylation. Genome Res 2015;25(06):907–917

17 Kang EY, Martin LJ, Mangul S, et al. Discovering single nucleotide polymorphisms regulating human gene expression using allele specific expression from RNA-seq data. Genetics 2016;204(03):1057–1064

18 Olsson Lindvall M, Hansson L, Klasson S, Davila Lopez M, Jern C, Stanne TM. Hemostatic genes exhibit a high degree of allele-specific regulation in liver. Thromb Haemost 2019;119(07):1072–1083

19 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9(04):357–359

20 Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics 2016;32(20):3207–3209

21 GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348(6235):648–660

22 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

23 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

24 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

25 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

26 Olsson M, Stanne TM, Pedersen A, et al. Genome-wide analysis of genetic determinants of circulating factor VII-activating protease (FSAP) activity. J Thromb Haemost 2018;16(10):2024–2034

27 Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. Genome Biol 2016;17:61

28 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

29 Kundaje A, Meuleman W, Ernst J, et al; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. Nature 2015;518(7539):317–330

30 Teichert M, Eijgelsheim M, Uitterlinden AG, et al. Dependency of phenprocoumon dosage on polymorphisms in the VKORC1, CYP2C9, and CYP4F2 genes. Pharmacogenet Genomics 2011;21 (01):26–34

31 Peden JFHJ, Saleheen D, Chambers JC, et al; Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet 2011;43(04):339–344

32 Ikram MA, Seshadri S, Bis JC, et al. Genome-wide association studies of stroke. N Engl J Med 2009;360(17):1718–1728

33 Do C, Shearer A, Suzuki M, et al. Genetic-epigenetic interactions in cis: a major focus in the post-GWAS era. Genome Biol 2017;18(01):120

34 Kulis M, Heath S, Bibikova M, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lympho-cytic leukemia. Nat Genet 2012;44(11):1236–1242

35 Varley KE, Gertz J, Bowling KM, et al. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res 2013;23 (03):555–567

36 Hannon E, Spiers H, Viana J, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat Neurosci 2016;19(01):48–54

37 Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. eLife 2013;2:e00523

38 Gutierrez-Arcelus M, Ongen H, Lappalainen T, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLoS Genet 2015;11(01):e1004958

39 Fu J, Wolfs MG, Deelen P, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet 2012;8(01):e1002431

40 Bettscheider M, Kuczynska A, Almeida O, Spengler D. Optimized analysis of DNA methylation and gene expression from small, anatomically-defined areas of the brain. J Vis Exp 2012;(65):e3938

41 Gamazon ER, Badner JA, Cheng L, et al. Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. Mol Psychiatry 2013;18(03):340–346

42 Do C, Lang CF, Lin J, et al. Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. Am J Hum Genet 2016;98(05):934–955

43 Volkov P, Olsson AH, Gillberg L, et al. A genome-wide mQTL analysis in human adipose tissue identifies genetic variants associated with DNA methylation, gene expression and metabolic traits. PLoS One 2016;11(06):e0157776

44 Pirmohamed M, Burnside G, Eriksson N, et al; EU-PACT Group. A randomized trial of genotype-guided dosing of warfarin. N Engl J Med 2013;369(24):2294–2303

45 Kimmel SE, French B, Kasner SE, et al; COAG Investigators. A pharmacogenetic versus a clinical algorithm for warfarin dosing. N Engl J Med 2013;369(24):2283–2293

46 Cullell N, Carrera C, Muiño E, Torres N, Krupinski J, Fernandez-Cadenas I. Pharmacogenetic studies with oral anticoagulants. Genome-wide association studies in vitamin K antagonist and direct oral anticoagulants. Oncotarget 2018;9(49):29238–29258

47 Luo Z, Liu R, Sun B, et al. Identification of gene modules associated with warfarin dosage by a genome-wide DNA methylation study. Pharmazie 2018;73(05):288–293

48 Gallego-Fabrega C, Carrera C, Reny JL, et al. PPM1A methylation is associated with vascular recurrence in aspirin-treated patients. Stroke 2016;47(07):1926–1929

49 Cotto KC, Wagner AH, Feng YY, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Res 2018;46(D1):D1068–D1073

50 Yang Y, Scott SA. DNA methylation profiling using long-read single molecule real-time bisulfite sequencing (SMRT-BS). Methods Mol Biol 2017;1654:125–134