

On the role of genetic variation and epigenetics in hemostatic gene regulation

Martina Olsson Lindvall

Department of Laboratory Medicine
Institute of Biomedicine
Sahlgrenska Academy, University of Gothenburg



UNIVERSITY OF GOTHENBURG

Gothenburg 2020

Cover illustration by Sofia Klasson

On the role of genetic variation and epigenetics in hemostatic gene regulation

© Martina Olsson Lindvall 2020

martina.olsson@gu.se

ISBN 978-91-7033-780-4 (PRINT)

ISBN 978-91-7833-781-1 (PDF)

Printed in Gothenburg, Sweden 2020

Printed by BrandFactory

Hur smått blir allting som fått ett svar!

Det stora är det som står olöst kvar,

När tanken svindlande stannat.

Bo Bergman, ur *Drömmarena*

ABSTRACT

Many genetic variants have been identified to associate with circulating levels of hemostatic proteins and with thrombotic or hemorrhagic disorders. However, the underlying molecular mechanisms remain largely unknown.

The overall aim of this thesis was to study how genetic variation and epigenetic mechanisms influence the regulation of hemostatic gene expression. The specific aims were to investigate epigenetic mechanisms regulating tissue-type plasminogen activator (t-PA) gene expression in the human brain (Paper I); to identify *cis*-acting variants involved in hemostatic gene regulation in liver (Papers II and III); and to investigate whether DNA methylation patterns in hemostatic genes in blood can reliably predict those in liver (Paper IV).

In Paper I, human astrocytes and neurons were treated with histone deacetylase (HDAC) inhibitors. Protein and mRNA levels of t-PA were measured using ELISA and real-time qPCR, respectively. Histone modifications were assayed with chromatin immunoprecipitation, and DNA methylation analysis of the t-PA promoter was performed by bisulfite sequencing.

In Papers II-IV, liver tissue and blood samples were collected from patients undergoing liver surgery and targeted DNA-, RNA- and methylation sequencing was performed for 35 hemostatic genes with predominant expression in the liver. These data were used in Papers II and III to performed allele-specific analyses of mRNA expression (ASE) and DNA methylation (ASM) in liver. In Paper IV, the extent to which blood can be used as a surrogate for DNA methylation of hemostatic genes in the liver was investigated.

In Paper I, cell treatments with HDAC inhibitors resulted in an increase in t-PA mRNA and protein expression, and in a significant increase in histone H3 acetylation. DNA methylation analysis revealed that the t-PA promoter was hypomethylated in neurons, astrocytes, and in post-mortem brain tissue, which indicates active transcription. In Paper II, ASE was identified in 60% of the hemostatic genes studied and 14 novel genotype-expression associations were discovered. In Paper III, a detailed DNA methylation map of the targeted hemostatic genes in liver was created, and novel associations between SNPs and DNA methylation were identified. The analyses performed in Paper IV showed that the correlation of hemostatic gene methylation between liver and blood was generally low. However, about 3% of the investigated CpGs had methylation levels that were significantly correlated between the two tissues,

and for these, blood may potentially be used as a surrogate tissue to detect liver methylation.

Taken together, these findings highlight the importance of integrating genetic, epigenetic, and expression analyses in the relevant tissue, and demonstrate that this approach can contribute to new insights into the biological processes affecting hemostasis and thrombosis.

Keywords:

Hemostasis, tissue-type plasminogen activator, genetics, epigenetics, DNA methylation

POPULÄRVETENSKAPLIG SAMMANFATTNING

Hjärtinfarkt och hjärninfarkt (även kallad ischemisk stroke) är bland de vanligaste dödsorsakerna i Sverige i dag, och hjärninfarkt är den enskilt största orsaken till bestående funktionsnedsättning hos vuxna. En infarkt uppstår oftast till följd av en blodpropp i ett tillförande kärl vilket leder till syrebrist med efterföljande vävnadsskador. Blodproppar kan även uppstå i de kärl som återför blodet i form av djup ventrombos eller lungemboli. En blodpropp kan bland annat uppstå till följd av rubbningar i blodets levringsprocess, även benämnd hemostas. Hemostasen är en essentiell process i kroppen vars uppgift är att stoppa blödningar och samtidigt undvika blodproppar. Systemet är komplext och involverar många olika typer av proteiner. Två centrala delar i hemostasen är *koagulering*, vilket innebär att blodet levrar sig för att täppa till och slutligen reparera en kärlskada, och *fibrinolys*, som är kroppens eget system för att lösa upp blodproppar. Det är av yttersta vikt att balansen i alla system som är involverade i hemostasen är noggrant reglerad. Blir balansen förskjuten i ena riktningen ökar risken för blodproppar och infarktsjukdomar, och blir balansen förskjuten i motsatt riktning ökar risken för blödningar.

Genom åren har flera genvarianter kunnat kopplas till förändrade nivåer av koagulationsproteiner och fibrinolysfaktorer, och även till en ökad risk för blodpropp, infarktsjukdomar eller blödningar. Genom vilka molekylära mekanismer dessa genvarianter har sin effekt är dock till stor del fortfarande okänt. I de delarbeten som presenteras i den här avhandlingen har vi därför valt att studera hur vissa genetiska och molekylärgenetiska (epigenetiska) mekanismer kan påverka uttrycket av flera gener som styr hemostasen.

Vi har främst studerat två olika typer av epigenetiska mekanismer: DNA-metylering och histonmodifieringar. DNA-metylering är en kemisk modifiering av det genetiska materialet, DNA, som innebär att en metylgrupp adderas till en eller flera positioner i DNA-molekylen. Histonmodifieringar innebär olika typer av förändringar på de proteiner, histoner, som organiserar DNA-molekylerna i cellkärnan. Båda dessa mekanismer har tidigare visats kunna påverka geners uttrycksnivåer. I delarbete I fann vi att histonmodifieringar i genen som kodar för den viktiga fibrinolysfaktorn vävnadsplasminogenaktivator, förkortat t-PA, var förknippade med uttrycket av denna gen i hjärnan, och att denna process sannolikt även involverar DNA-metylering. Genen som kodar för t-PA är normalt högt uttryckt i hjärnan där t-PA-proteinet, utöver hemostasen, även är involverat i flera andra processer såsom inlärning och vid återhämtning av funktioner efter hjärninfarkt.

Vid epigenetiska studier, såväl som vid studier av genuttryck, är det av stor vikt att undersöka just den vävnad som genen eller generna av intresse är uttryckta i. Detta på grund av att både epigenetiska mekanismer och genuttryck kan variera stort mellan olika vävnader i kroppen. Många av de proteiner som styr hemostasen uttrycks och produceras till största del i levern, varifrån de sedan utsöndras till blodets cirkulationssystem. I delarbetena II-IV studerade vi därför 35 hemostasgener som framförallt uttrycks i levern. I delarbete II studerade vi hur genvarianter påverkar uttrycket av dessa gener, och identifierade ett flertal tidigare okända kopplingar mellan genetiska varianter och uttrycket av hemostasgener. I delarbete III presenterade vi för första gången en detaljerad bild över metyleringsmönstret i dessa 35 hemostasgener i levervävnad. Vi identifierade nya kopplingar mellan genetiska varianter och graden av DNA-metylering i dessa gener i levern. I delarbete IV undersökte vi i vilken grad DNA-metyleringen i de 35 hemostasgenerna i blod kan spegla metyleringsmönstret i levern. Vi fann att för den absoluta majoriteten av metyleringspositioner förelåg ingen korrelation mellan metyleringsgraden i blod och lever inom samma individ. Resultaten visar således att blod bara kan användas för att undersöka metyleringsmönster i lever för ett fåtal bestämda positioner i genomet.

Sammantaget visar resultaten från den här avhandlingen att både genetiska och epigenetiska mekanismer har betydelse för regleringen av uttrycket av ett relativt stort antal gener som styr hemostasen. Genom modern sekvenseringsteknik som genererar stora datamängder och avancerade analyser har vi lyckats skapa en detaljerad karta över dessa samband, och identifierat nya kopplingar mellan genetik, epigenetik och genuttryck. Slutligen visar resultaten att det för epigenetiska studier är av stor vikt att undersöka dessa mekanismer i den mest relevanta vävnaden.

LIST OF PAPERS

This thesis is based on the following papers, referred to in the text by their Roman numerals.

- I. Olsson M, Hultman K, Dunoyer-Geindre S, Curtis MA, Faull RLM, Kruithof EKO, Jern C. Epigenetic regulation of tissue-type plasminogen activator in human brain tissue and brain-derived cells. *Gene Regulation and Systems Biology* 2016;10:9-13.
- II. 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. *Thrombosis and Haemostasis* 2019;119:1072-1083.
- III. Olsson Lindvall M, Davila Lopez M, Klasson S, Hansson L, Nilsson S, Stanne TM**, Jern C**. A comprehensive sequencing-based analysis of allelic methylation patterns in hemostatic genes in human liver. *Thrombosis and Haemostasis* 2020; *Epub ahead of print*.
- IV. Olsson Lindvall M*, Angerfors A*, Andersson B, Nilsson S, Davila Lopez M, Hansson L, Stanne TM**, Jern C**. Comparison of DNA methylation profiles of hemostatic genes between liver tissue and peripheral blood within individuals. *In manuscript*.

*These authors contributed equally to this work.

**These authors jointly supervised this work.

All papers are appended at the end of this thesis. Reprints were made with permission from the publishers.

CONTENTS

ABBREVIATIONS	IV
INTRODUCTION	1
Basic genetics	1
The central dogma	2
Genetic variation	3
Epigenetics	4
DNA methylation	6
Histone modifications	7
Overview of methods for studying the DNA sequence	8
Historical aspects	8
Sanger sequencing	8
Polymerase chain reaction	9
Pyrosequencing	9
Next generation sequencing	9
Sequencing of the human genome	10
Genome-wide association studies	11
Quantitative trait loci and allele-specific approaches	12
Hemostasis	13
Primary hemostasis	13
Secondary hemostasis	13
Fibrinolysis	14
Liver and hemostatic gene regulation	15
Genetic variation and circulating hemostatic proteins	15
Epigenetics and hemostasis	16
Role of hemostatic proteins in arterial and venous thrombosis	16
Role of hemostatic proteins in the central nervous system	17
AIM OF THE THESIS	18
SUBJECTS AND METHODS	19
Material	19
Cell culture	19
Cell culture and treatments of human neurons and astrocytes (Paper I)	20
Human brain tissue	21
Human brain tissue collection (Paper I)	21
Human liver	22
Human liver tissue and blood sample collection (Papers II-IV)	22
mRNA quantification	23
Quantitative polymerase chain reaction	23
RNA sequencing	24
Analysis of t-PA mRNA expression using qPCR (Paper I and II)	24
Analysis of hemostatic mRNA expression (Papers II and III)	25
Protein analyses	25
Analysis of t-PA protein concentrations (Paper I)	25
Chromatin immunoprecipitation assay	26

Analysis of histone modifications in the t-PA promoter (Paper I)	26
Methylation sequencing of the t-PA promoter (Paper I)	27
Design of a targeted sequencing panel for hemostatic genes (Papers II-IV)	28
Massive parallel sequencing	30
DNA sequencing of hemostatic genes (Papers II-III)	31
RNA sequencing	31
RNA sequencing of hemostatic genes (Papers II-III)	31
Methylation sequencing	32
Methylation sequencing of hemostatic genes (Papers III-IV)	33
Sequence data quantitation and downstream analyses	33
Allele-specific expression analysis	33
Analysis of allele-specific expression of hemostatic genes (Paper II)	34
Allele-specific DNA methylation analysis	34
Analysis of allele-specific DNA methylation in hemostatic genes (Paper III)	34
Databases and tools	35
Statistical analyses	36
Ethical approvals and considerations	37
RESULTS AND DISCUSSION	38
Regulation of t-PA in the human brain and brain-derived cells (Paper I)	38
Sequencing output and evaluation of the targeted design (Papers II-IV)	40
Hemostatic genes exhibit allele-specific expression in human liver tissue (Paper II)	42
Allele-specific DNA methylation of hemostatic genes in human liver (Paper III)	43
Overlap between allele-specific expression and DNA methylation (Papers II-III)	45
Correlation of DNA methylation levels between liver and blood (Paper IV)	45
CONCLUSIONS TO GIVEN AIMS	47
CONCLUDING REMARKS AND FUTURE PERSPECTIVES	48
ACKNOWLEDGEMENTS - TACK	50
REFERENCES	52

ABBREVIATIONS

5mC	5-methylcytosine
A	adenine
A1AT	alfa-1-antitrypsin
AGM	astrocyte growth medium
AS	alternative splicing
ASE	allele-specific expression
ASE-SNP	ASE associated SNP
ASM	allele-specific methylation
ASM-SNP	ASM associated SNP
BBB	blood brain barrier
BDNF	brain derived neurotrophic factor
bp	base pairs
C	cytosine
C4BPB	complement component 4 binding protein
CAD	coronary artery disease
cDNA	complementary DNA
CGI	CpG island
CHD	coronary heart disease
ChIP	chromatin immunoprecipitation
CNS	central nervous system
CNV	copy number variation
CPB2	carboxypeptidase B2
CpG	cytosine-phosphate-guanine
dbSNP	Single Nucleotide Polymorphism Database
ddNTP	dideoxynucleotide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	deoxynucleotide
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EMBL	European Molecular Biology Laboratory
EMBL-EBI	EMBL European Bioinformatics Institute
ENCODE	Encyclopedia of DNA Elements
eQTL	expression quantitative trait loci
EWAS	epigenome-wide association study
FDR	false discovery rate
FGB	fibrinogen beta chain

FIV	factor IV
FIX	factor IX
FPKM	fragments per kilobase million
FSAP	FVII activating protease
FV	factor V
FVII	factor VII
FVIII	factor VIII
FX	factor X
FXI	factor XI
G	guanine
gDNA	genomic DNA
GP	glycoprotein
GRS	genetic risk score
GTE _x	Genotype-Tissue Expression
GWAS	genome-wide association study
HAT	histone acetyltransferase
HDAC	histone deacetylase
HMT	histone methyltransferase
HNF4	hepatocyte nuclear factor 4
HUGO	Human Genome Organisation
IC ₅₀	half maximal inhibitory concentration
kb	kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
KNG1	kininogen 1
LD	linkage disequilibrium
L-LTP	late phase of long-term potentiation
lncRNA	long non-coding RNA
LRP	low-density lipoprotein receptor-related protein
Mb	megabase
MI	myocardial infarction
miRNA	micro RNA
mQTL	methylation quantitative trait loci
mRNA	messenger RNA
MSigDB	Molecular Signatures Database
NCBI	National Center for Biotechnology Information
NF-I	nuclear Factor I
NGI	National Genomics Infrastructure
NGS	next-generation sequencing
NHGRI	National Human Genome Research Institute
M	million
NMDAR	N-methyl-D-aspartic acid receptor
NOAC	non-vitamin K antagonist oral anticoagulants

PAI-1	plasminogen activator inhibitor type 1
PAR	protease-activated receptors
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PIC	protein inhibitor cocktail
PMSF	phenylmethylsulfonyl fluoride
pQTL	protein quantitative trait loci
PROZ	protein Z
qPCR	quantitative polymerase chain reaction
QTL	quantitative trait loci
RCF	relative centrifugal force
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RPK	reads per kilobase
RPKM	reads per kilobase million
RPM	reads per million
RXRA	retinoid X receptor alpha
SAHLSIS	Sahlgrenska Academy Study on Ischemic Stroke
SERPINA5	Serpin family A member 5
SERPINF2	Serpin family F member 2
SNP	single nucleotide polymorphism
T	thymine
TAFI	thrombin-activatable fibrinolysis inhibitor
TF	tissue factor
TFPI	tissue factor pathway inhibitor
t-PA	tissue-type plasminogen activator
TPM	transcripts per kilobase million
TSA	trichostatin A
TSS	transcription start site
TXA2	thromboxane A2
U	uracil
u-PA	urokinase-type plasminogen activator
UCSC	University of California, Santa Cruz
VKORC1	vitamin K epoxide reductase complex subunit 1
VTE	venous thromboembolism
vWF	von Willebrand factor
WGS	whole genome sequencing
ZPI	protein Z-dependent protease inhibitor

INTRODUCTION

Basic genetics

Within most of the different cell types in the human body lies a cell nucleus. The nucleus is the core organelle in which the DNA (deoxyribonucleic acid) resides. Human DNA is arranged in two copies of 23 large molecules called chromosomes. One member of each chromosome pair is inherited from the mother, and the other one from the father. Chromosomes 1–22 are autosomal, and all genes and elements located on these thus occur in two versions in each cell - the maternal allele and the paternal allele. The last pair are the sex chromosomes, which contains two X chromosomes for females and one X chromosome and one Y chromosome for males.¹

A human DNA molecule is organized as two complementary spiral-shaped strands. Each strand consists of a sugar-phosphate backbone and the four nucleotide bases: cytosine (C), thymine (T), adenine (A) and guanine (G). The two strands are connected by hydrogen bonds, where C always pairs with G and T always pairs with A, into a structure known as a double helix.² The order and combination of these four nucleotides comprises the genetic code, containing information for all cellular processes.

The double helices are folded in several layers by various DNA binding proteins, histones being the most abundant type, and organized into nucleoprotein complexes, i.e. the chromatin. Cellular processes are mainly carried out by ribonucleic acid (RNA) molecules, proteins, and peptides. The genes in the DNA contain instructions for how these elements are constructed, and this is regulated by reorganization of chromatin packaging. Actively transcribed genes tend to be located in loosely packed chromatin regions known as euchromatin, whereas inactive genes are located in the more densely packed heterochromatin. The reorganization of chromatin and the resulting gene regulation is an ongoing, dynamic process that varies between cell-types, time points, and in response to both internal and external factors.^{3,4}

The central dogma

Three main mechanisms are involved in the processing and interpretation of the genetic code: 1) replication, 2) transcription, and 3) translation. Replication occurs when two identical copies of a DNA molecule are produced, using the original molecule as a template. This takes place when cells undergo mitosis, i.e. divide so that two daughter cells are formed from a single parent cell. Transcription is the process in which the information in genes is copied and stored in molecular templates which will be used to produce proteins. These templates are single-stranded nucleic acids called messenger RNA (mRNA). The mRNA consists of three of the same bases as DNA: C, A and G, but with uracil (U) being the fourth base instead of T. The mRNA molecule is a complimentary copy of the DNA sequence comprising the gene in question. The last step, translation, is the process in which the mRNA templates are read and the stored information is interpreted. The mRNA is decoded in the translational machinery, the ribosome, which links amino acids together in the order specified by the genetic code. The mRNA is read three nucleotides at a time. Each group of trinucleotides, or “codons”, specify which amino acid will be added next during protein synthesis.¹ The principles of the central dogma are illustrated in **Figure 1**.

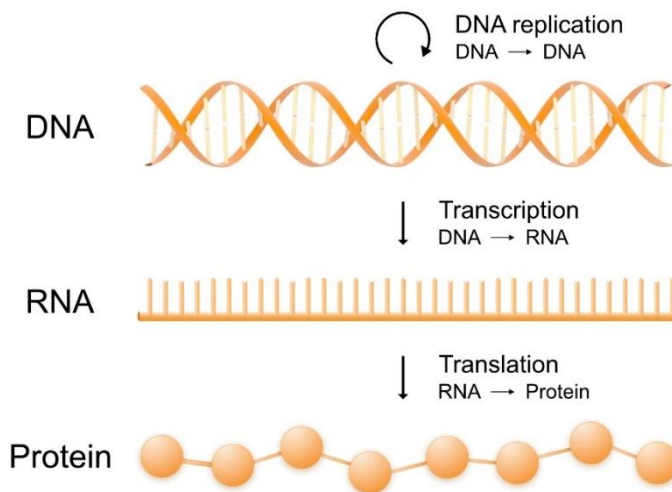


Figure 1. Principles of the central dogma. DNA is copied during cell division (replication). RNA is transcribed into a complimentary copy of the DNA template (transcription). Amino acids are assembled into a protein based on the genetic code stored in the RNA (translation).

Genetic variation

The genetic material in the nucleus of each human cell consists of around three billion nucleotide base pairs.^{5,6} Although the genetic differences between individuals are very small, they are in no regard negligible. The underlying source of genetic diversity is acquired and inherited alterations in the genome. There are several types of both small sequence variations, such as single nucleotide polymorphisms (SNPs; **Figure 2**), and larger structural variations, such as chromosomal rearrangements or copy number variation (CNV).⁷ All differences between individuals are a consequence of these types of diversity in combination with environmental factors.

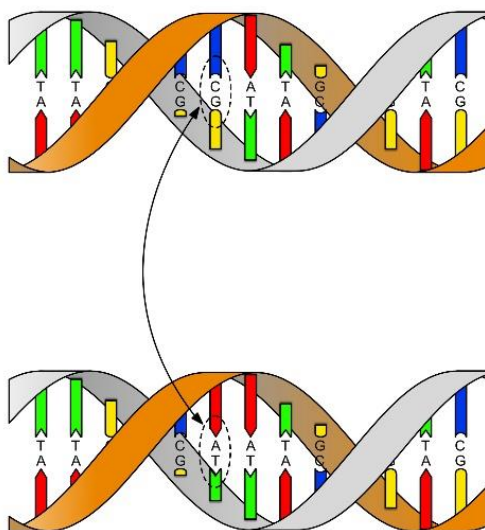


Figure 2. The genomic sequence may differ between individuals due to a substitution of a single nucleotide base, a so called single nucleotide polymorphism (SNP), at a specific location in the genome.

A SNP is a substitution of a single nucleotide at a specific locus in the genome. If a carrier has the same nucleotide on both alleles, it is classified as homozygous for that position, and if it has two different nucleotides, it is heterozygous. Depending on where the SNP is located, it can have different consequences. SNPs in coding regions are either so called synonymous or non-synonymous. Synonymous SNPs do not affect the protein sequence and represent the vast majority of the coding SNPs. Non-synonymous SNPs either cause a change in the amino acid sequence (missense), potentially resulting in defective protein isoforms, or introduce a premature stop codon (nonsense) which results in a truncated protein. SNPs that fall outside of coding regions can also have functional effects, for example by affecting transcription factor binding affinity or by affecting splicing or mRNA stability.⁸

The alleles of SNPs located in proximity to each other in the genome are unlikely to be separated during recombination, and are thus highly correlated. Clusters of linked SNPs are referred to as haplotype groups (haplogroups), and are useful in the sense that a few SNPs can be used in order to determine the alleles of the remaining SNPs within the same haplogroup, within one individual. Thus, haplogroup information is essential in the search for genetic associations to diseases and other traits, as it reduces the number of SNPs required for genome-wide examinations.⁹

Epigenetics

Epigenetics is the study of variations that do not involve changes in the DNA sequence. These mechanisms are essential during developmental processes, cell differentiation, X chromosome inactivation and imprinting, and cell- and tissue-specific gene expression.^{10,11} The epigenetic state of an individual's genome varies between tissues. It also changes during developmental processes and aging, and can be influenced by environmental factors such as diet, life style factors, and various diseases.¹² Epigenetics is thus an important field in modern medicine as it may help to combine and explain the relationship between an individual's genetic background, environmental factors, and disease. Two major epigenetic mechanisms include DNA methylation and histone modifications, and both impact chromatin remodeling and gene regulation by diverse processes. An illustration of these main epigenetic mechanisms is shown in **Figure 3**.

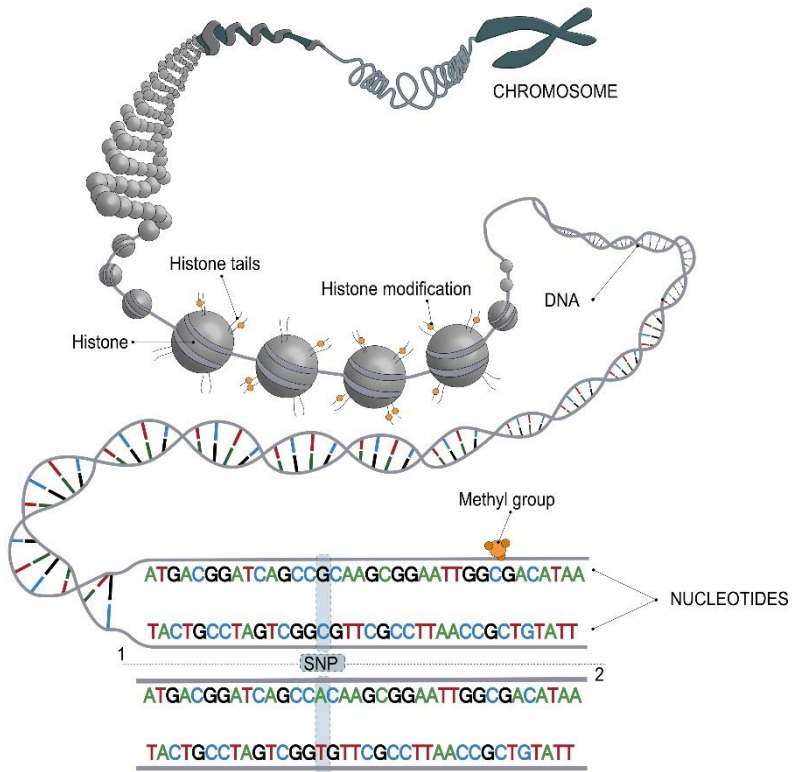


Figure 3. Schematic view of epigenetic mechanisms. DNA methylation involves the addition of a methyl group to a cytosine nucleotide. Post-translational modifications occur on histone tails which alter their interaction with other nuclear proteins and with DNA. Epigenetic mechanisms are responsible for the organisation of chromatin in the cell nucleus.

DNA methylation

DNA methylation is the most stable epigenetic mark, and has been implicated in several complex human traits and diseases. Methylation of DNA is the result of amination of cytosines, which involves the addition of a methyl group to the fifth carbon of the pyrimidine ring, converting the cytosine into a 5-methylcytosine (5mC). Methylation can be maintained during cell division and inherited by the daughter cells. This is managed by DNA methyltransferase I (DNMT1) which, following replication, recognizes hemi-methylated DNA sequences and methylates the cytosines on the newly synthesized unmethylated daughter strand using the old methylated strand as a template.¹¹ Methylation usually occurs on cytosines in a CG sequence context, also known as CpG dinucleotides. Over time, methylated cytosines in the CpG context are prone to undergo spontaneous deamination. Deamination of methylated cytosines will result in conversion of the 5mC to a T base, resulting in a T-G base pair mismatch. DNA repair mechanisms will attempt to ameliorate this by either changing the T back to a C, thereby restoring the original C-G base pairing, or by changing the G base to an A, resulting in T-A base pairing. The latter case results in a sequence variant in the DNA. As discussed above, sequence variants can have deleterious effects on cell function, and for this reason, CpG dinucleotides are depleted in most parts of the genome, compared to other nucleotide sequence combinations.¹³ However, certain genomic regions with a high density of CpGs, known as CpG islands (CGI), are also present.¹⁴ These regions are often located in DNA regulatory elements, such as transcription factor binding sites in promoters and enhancers, and CGI methylation in these contexts has been shown to repress gene transcription.¹⁵ More recently, however, DNA methylation within gene bodies (i.e. exons and introns) has been reported to also be prevalent and to enhance gene expression.¹⁶⁻¹⁸ Thus, the relationship between DNA methylation and gene expression is highly complex and not fully understood.

In addition to this, DNA methylation can also be influenced by genetic variants. In diploid genomes, the two alleles can exhibit different methylation patterns, known as allele-specific DNA methylation (ASM). ASM is well established in imprinted genes and X chromosome inactivation¹⁹, but has recently also been identified in autosomes. This type of allelic asymmetry is thought to be mostly accounted for by *cis*-acting regulatory SNPs.^{20,21} Therefore, characterizing the relationship between genetic variation and DNA methylation could provide insights into mechanisms regulating phenotypic diversity, such as susceptibility to certain diseases, and response to drugs and environmental agents.

Histone modifications

Histone modifications are, in contrast to DNA methylation, readily reversible and are usually not maintained during cell division.²² Histones are the DNA binding proteins mainly responsible for the packaging of DNA into chromosomes. There are four core types of histones: H2A, H2B, H3, and H4. They occur as homodimers under normal circumstances. The DNA is wrapped around a histone octamer formed by one dimer of each of the four types, into the so-called nucleosomes structure, and these are separated by approximately 50 bp of DNA. This most basic formation is known as the 10 nm fiber or “beads-on-a-string formation”, referring to the resemblance of the nucleosomes as beads arranged along a string. This formation is then arranged in several higher-order structures, eventually forming the chromatin (**Figure 3**).¹

The N-terminal of histones consists of the histone “tail”. Amino acid residues in histone tails are exposed to various post-translational modifications such as acetylation, methylation, and phosphorylation, among others, which alter their interaction with other nuclear proteins and with DNA. The impact of histones on gene transcription is governed by chromatin remodeling, and mediated by a combination of different types of these modifications. These patterns are usually complex, but as a general rule of interpretation, methylation of histones is considered to induce gene silencing, while acetylation is considered to induce gene activation.^{23,24}

Histone acetylation is generated and maintained by enzymes known as histone acetyltransferases (HAT), and are removed by histone deacetylases (HDAC). There is an interplay between histone modifications and DNA methylation, and they often act together to recruit various chromatin remodeling complexes. For example, methylated DNA can bind various methyl binding proteins which, in turn, are able to recruit both HDACs and histone methyltransferases (HMTs). HDACs and HMTs can deacetylate and methylate nearby histones, and thereby further promote gene repression. Methylated histones also have the ability, via binding of various chromodomain proteins, to recruit additional DNMTs, which further contributes to DNA methylation and gene silencing.^{22,25}

Overview of methods for studying the DNA sequence

Historical aspects

The structure of the DNA double helix was solved by James Watson and Francis Crick as early as 1953², largely based on crystallographic data produced by Rosalind Franklin²⁶ and Maurice Wilkins²⁷. Although the structure had been described in detail, the technology needed to be able to “read” the genetic sequence of DNA from living organisms had not yet been developed. The development of sequencing methods was initially focused on single-stranded RNA, which is less complex than DNA, and the first methods started to appear in the mid-1960s.²⁸⁻³¹ The first primitive DNA sequencing methods were based on these technologies and were initially performed on DNA with “overhanging” 5’ ends from bacteriophages. The overhanging ends facilitated the use of DNA polymerase to insert radioactively labeled nucleotides which were supplied one at a time while monitoring the sequence incorporation. In the early 1970s, this method was adapted with the use of oligonucleotides to “prime” DNA polymerase (thereby escaping the need for overhanging ends) and could be implemented on all types of DNA.³²⁻³⁴ Around this time, advances in electrophoresis (i.e. the separation of particles using an electrical charge) were also developed, improving the efficiency and resolution greatly.^{35,36}

Sanger sequencing

In 1977, Fred Sanger described a novel DNA sequencing approach, the chain-termination technique, developed at his laboratory.³⁷ This was a major breakthrough in sequencing technology, and it largely resembles the methods used today. This approach is based on the use of the analog nucleotides deoxynucleotides (dNTP) and radioactively labeled, dideoxynucleotides (ddNTPs). ddNTP lacks the hydroxyl group required for DNA extension by DNA polymerase. When the polymerase synthesizes the complementary strand, ddNTPs are occasionally incorporated instead of dNTPs which results in termination of the reaction, yielding DNA fragments of various lengths. Four reactions, each containing one of the four ddNTP bases (ddATP, ddTTP, ddCTP or ddGTP), are performed in parallel and run on a polyacrylamide gel by electrophoresis. Autoradiography can then be used to determine the nucleotide sequence as the fragments are separated by size. Over the years, further improvements have been made to this method and today it is performed using fluorescence labeling rather than radioactivity. The four terminating

ddNTPs are labelled with fluorophores of different wavelengths, and the DNA sequence can then be determined by laser detection of the fluorophores when the fragments are separated. This method is still commonly referred to as “Sanger sequencing”, after the developer.

Polymerase chain reaction

In 1983, biochemist Kari Mullis developed a new method for DNA amplification, the polymerase chain reaction (PCR).³⁸ The technique is based on thermal cycling and temperature-dependent reactions. In the reaction, the double-stranded template DNA is first heated to the point of denaturation. The temperature is then lowered to allow binding of oligonucleotide primers to the two separated DNA strands, and a DNA polymerase initiates enzymatic assembly of two new double strands by incorporating nucleotides available in the reaction mixture. This cycle is then repeated, generating an exponential amplification of the original DNA template. The PCR method is now widely used both for medical diagnostic purposes and in research, and it is one of the fundamental steps in modern sequencing techniques.

Pyrosequencing

In 1993, yet another sequencing method based on luminescence detection using the firefly luciferase enzyme was introduced.^{39,40} The reaction relies on the release of pyrophosphate that occurs when a dNTP is incorporated by the polymerase during DNA synthesis, and thus became known as pyrosequencing. As with Sanger sequencing, each dNTP is added to the reaction one at a time. When the nucleotide is incorporated and pyrophosphate is released, the firefly luciferase acts on its substrate luciferin which then produces detectible light. The intensity of the light is proportional to the number of dNTPs incorporated, thus enabling determination of the number of dNTPs incorporated in each PCR cycle.

Next generation sequencing

Next generation sequencing (NGS), also denoted massive parallel sequencing, first appeared around 2005.⁴¹ However, the speed, the number of samples and the amount of DNA that can be sequenced per run, and thus the amount of output data generated per unit of time has increased exceptionally since the development of this new sequencing technology.⁴² First, 5' and 3' adapters are

added to each DNA fragment in the sequence library. The DNA fragments are then loaded into a flow cell and captured by millions of surface-bound oligonucleotides, complementary to the library adapters. Each bound DNA fragment is amplified in clonal clusters through bridge amplification. Primers attach to the forward or reverse strand and DNA polymerase adds the fluorescently labeled nucleotides one by one. Each of the four bases has a unique emission which is recorded after each amplification round. NGS also allows for multiplexing of several samples in one single run. This is achieved by adding sample-specific oligonucleotide indexes, or “barcodes”, to each DNA fragment before sequencing. The indexes are sequenced during the reaction, and the generated data can be separated based on the sample-specific index sequence.⁴³ With these platforms, gigabases of sequence reads can be generated simultaneously in one single instrument run⁴⁴, and the cost for DNA sequencing is constantly declining⁴⁵. Not surprisingly, this new technology has thus had a fundamental impact on genetic research.⁴⁶

Sequencing of the human genome

In 1990, long before NGS had been invented, the Human Genome Project was launched with the aim to sequence the entire human genome and to identify all human genes. It was initiated by the US government, and performed as a large international research collaboration overseen by the Human Genome Organisation (HUGO).⁴⁷ In 1998, the private company Celera announced that they intended to launch a similar project purposed to proceed faster than the HUGO project and at a lower cost.⁶ Ten years after the HUGO project was initiated, the two leaders of the competing groups announced that a first draft sequence covering around 90% of the human genome was complete.^{6,13} The finalized genome sequence, covering approximately 99% of the genome, was considered complete first two years later, in 2003.⁵ The publicly funded HUGO project thus took thirteen years, and costed approximately 3 billion US dollars.⁴⁸ The private initiative by Celera took five years and costed around 100 million US dollars to complete.⁴⁹ However, the Celera project could draw on data that had already been made available by HUGO.

Shortly after the sequence of the human genome was announced, a number of large genome projects were initiated. One of the first was the International HapMap project⁵⁰, which was launched in 2002, with the objective to describe common genetic variation and to develop a haplotype map of the human genome. The last dataset from HapMap was released in 2010. It is based on DNA samples from ~1,200 individuals from a variety of human populations⁵¹, and still remains an important data source for the research community.

Another project with a similar ambition is the Encyclopedia of DNA Elements (ENCODE). It was launched by the National Human Genome Research Institute (NHGRI) in 2003, and intended as a follow-up to the HUGO project.⁵² The main focus is to describe the functional elements in the human genome. The ENCODE encyclopedia now accommodates information on gene expression, promoter activity, transcription factor binding sites, open chromatin and chromatin structure, histone modifications, DNA methylation, and much more.

The 1,000 genomes project was an international collaborative project conducted between 2008 and 2015, with the aim of sequencing the genomes of at least 1,000 individuals using the newly developed sequencing technologies.⁵³ The final dataset consists of sequencing data from around 2,500 individuals, and is the most detailed catalogue of common and rare human genetic variation today.⁵⁴

One of the largest whole genome sequencing (WGS) projects is the 100,000 Genomes Project.⁵⁵ It was initiated in the United Kingdom in 2012 with the ambition to sequence 100,000 human genomes from patients in the National Health Service, affected by rare diseases and cancer; a goal they met in 2018.⁵⁶

Genome-wide association studies

The sequencing efforts described above, such as the HapMap project and the 1,000 Genomes Project, have paved the way for genome-wide association studies (GWASs). Knowledge of the haplotype structure and linkage disequilibrium (LD) between genetic variants has been utilized to develop microarray chips for genotyping. Genotyping using these SNP arrays includes hybridization of sample DNA with allele-specific oligonucleotide probes, which are immobilized on a microarray chip. This allows direct determination of the alleles for each SNP included on the chip, but also for many other SNPs in the same haploblocks based on LD information. SNP arrays are used in GWASs to assay millions of genetic variants across the whole genome in a large number of subjects with different traits, or in subjects with or without the disease of interest. If one allele is more common in one of the groups at a genome-wide significance level, it is considered to be associated with the disease or trait in question.

Through GWASs, a large number of SNPs throughout the genome have now been associated with different quantitative traits such as height⁵⁷, blood pressure⁵⁸, and susceptibility to complex diseases such as coronary heart

disease⁵⁹, stroke⁶⁰, diabetes mellitus⁶¹, and Alzheimer's disease⁶². Furthermore, SNPs may not only affect disease susceptibility, but also the severity and outcomes of diseases⁶³, as well as the response to treatments and drugs⁶⁴. Genetic variants, such as SNPs, are thus of great interest in the current era of personalized drug development and precision medicine.

Quantitative trait loci and allele-specific approaches

Studies of the effects of SNPs on more intermediate traits are often performed by quantitative trait loci (QTL) studies. They are commonly conducted in a large sample with genotypic and phenotypic data.

Expression quantitative trait loci (eQTL) studies link variations in genotypes to mRNA expression across individuals and is one approach to elucidate whether genetic variants correlate to gene expression.⁶⁵ Several databases have been developed to collect eQTLs for different human tissues and cells, including the Genotype-Tissue Expression (GTEx) portal⁶⁶. Similar to eQTL, methylation QTL (mQTL) studies, which link variations in genotypes to DNA methylation, can be used to study correlations between genetic variants and the methylation level or pattern of specific CpGs, genes, or regions.⁶⁷

An alternative approach to the eQTL and mQTL studies is to perform analyses of allele-specific expression (ASE) and methylation (ASM). ASE is a quantitative phenomenon that results in a bias in the ratio of transcripts from the two alleles.⁶⁸ and has been successfully used to identify a few functionally important regulatory variants in hemostatic genes⁶⁹⁻⁷¹. ASE occurs when transcription from one allele is selectively silenced or enhanced, or when transcripts undergo selective post-transcriptional degradation (e.g., nonsense-mediated decay). By comparing expression of alleles within the same individual, each allele acts as an internal control for confounding factors (e.g., *trans*-acting effects and environmental confounders) that may alter the overall expression of that gene.⁷² ASM has been demonstrated to be widespread among autosomal non-imprinted genes.⁷³ This 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 between the two alleles.

Hemostasis

Hemostasis is the physiological process that regulates the intrinsic balance in order to maintain intravascular blood circulation and prevents blood loss after a vessel injury. The system is complex and involves three major processes. *Primary hemostasis* is the formation of a platelet plug, and it is initiated by adhesion of platelets to collagen fibers at the site of vessel injury. *Secondary hemostasis* involves blood clot formation and generation of the coagulation factor thrombin. *Fibrinolysis* is the endogenous breakdown of fibrin, which dissolves the clot when the injury is repaired. These processes are heavily regulated both by positive and negative feedback systems to prevent thrombotic events as well as excessive bleeding.

Primary hemostasis

Primary hemostasis involves three sequential steps: platelet adhesion, platelet activation, and platelet aggregation. Platelets, or thrombocytes, are circulating anucleate cells originating from the bone marrow. In the event of an injury to the blood vessel, platelets adhere to the damaged area through a series of events involving interaction of the platelet membrane receptor glycoprotein (GP) Ib-V-IX complex with the immobilized form of von Willebrand factor (vWF) on exposed extracellular matrix (ECM), and of the receptor GPVI on the platelet to exposed collagen at the site of injury. Platelet activation is initiated immediately following adhesion, in response to further interactions between cell surface receptors with factors in the exposed collagen. GPVI signaling also increases platelet secretion of thromboxane A₂ (TXA₂), which acts on the platelet's own cell surface receptors, and those of other platelets. These and other receptors trigger intracellular signaling pathways that convert different G protein coupled receptors to their active form, eventually initiating platelet aggregation, granule secretion of various coagulation factors and chemotactic agents, integrin activation, and cytoskeleton remodeling.^{74,75}

Secondary hemostasis

Secondary hemostasis consists of the coagulation cascade and involves an array of reactions catalyzed by serine proteases, eventually resulting in the cleavage of fibrinogen by thrombin, to generate fibrin. Vessel injury can initiate this cascade through the release of tissue factor (TF) from extravascular tissues. TF is then exposed to the circulating serine protease factor VII (FVII), which it activates. The two form a complex which activates FIX and FX. FX

is further activated by the active form of FIX (FIXa) and its cofactor FVIIIa. Finally, FXa activates prothrombin to generate thrombin, with the help of its cofactor FV. Thrombin is responsible for the cleavage of fibrinogen to generate insoluble fibrin which forms a crosslinked mesh at the site of an injury, in which red blood cells and platelets are trapped. Thrombin also activates platelets via cleavage of the protease-activated receptor (PAR) 1 and PAR4, which are responsible for the positive feedback mechanisms critical for clot generation.⁷⁶⁻⁷⁸

In contrast to this, thrombin is also a key player in the downregulation of coagulation. Here, thrombin binds to thrombomodulin on the surface of endothelial cells, thereby activating protein C which, in turn, cleaves and inactivates FVIIIa and FVa, together with the cofactor protein S.^{79,80} The coagulation cascade is further down-regulated by various serine protease inhibitors including antithrombin (inhibiting thrombin, FXa, FIXa and FXIa); heparin cofactor II (which also inhibits thrombin); protein Z-dependent protease inhibitor (an inhibitor of FXa); protein C inhibitor (the inhibitor of activated protein C); C1-inhibitor (which inhibits FXIa); and tissue factor pathway inhibitor (inhibiting FXa); and alpha-2-macroglobulin (which inhibits thrombin).⁸¹

Fibrinolysis

Dissolvement of blood clots is necessary to avoid thrombus formation in healthy vessels. Fibrinolysis is initiated when the serine protease tissue-type plasminogen activator (t-PA) binds to the surface of a fibrin clot. t-PA facilitates dissolution of fibrin-containing blood clots through cleavage of plasminogen into active plasmin which degrades fibrin.⁸² Fibrin potentiates the ability of t-PA to activate plasmin and thus systematic fibrinolysis does not generally occur in the absence of fibrin clots.⁸³ The fibrinolytic activity is also regulated by plasminogen activator inhibitor type 1 (PAI-1), which inhibits t-PA, and alpha-2-antiplasmin, which inhibits thrombin.⁸¹ In addition, thrombin-activatable fibrinolysis inhibitor (TAFI) reduces plasmin activity by modification of the fibrin residues necessary for the binding of plasmin to fibrin.⁸⁴

Liver and hemostatic gene regulation

While the hemostatic system mainly acts within the vascular compartment, a relatively large proportion of the hemostatic proteins originate from the liver⁸⁵, and it is well known that patients with liver disease often display coagulation abnormalities⁸⁶. Given this background, we have in Papers II-IV chosen to focus on the regulation of hemostatic genes specifically in liver tissue.

Genetic variation and circulating hemostatic proteins

Several mutations and SNPs have been identified that affect circulating levels or activity of hemostatic proteins. Hemophilia A and B are X-linked hereditary bleeding disorders caused by mutations in the genes encoding FVIII and FIX, respectively.⁸⁷ Similarly, autosomal mutations in the genes encoding FXI⁸⁸ and vWF⁸⁹ have been shown to cause bleeding disorders due to decreased levels of the respective proteins. There are also mutations in the genes encoding antithrombin⁹⁰, protein C⁹¹, and protein S⁹² that give rise to thrombotic disorders, and mutations in the genes encoding the fibrinogen subunits⁹³, prothrombin^{94,95} and FV^{96,97} have been associated with both bleeding and thrombotic events.

With regards to common variants, initial studies employed the candidate gene approach. Examples of SNPs within hemostatic genes that have been robustly associated with the respective plasma protein level using this approach include the common 4G/5G polymorphism in the promoter of the gene encoding PAI-1 (*SERPINE1*). The PAI-1 4G allele has consistently been associated with higher PAI-1 activity.^{71,98} Two variants in the β -fibrinogen promoter (-455G/A and -854G/A) have been consistently associated with fibrinogen levels⁹⁹⁻¹⁰², and two variants in the FVII gene (-40IG/T and -402G/A) have been linked to plasma FVII levels¹⁰³. Later, several genome-wide association studies on circulating hemostatic factors have been performed and identified loci associated with plasma concentrations or activity of circulating hemostatic proteins such as fibrinogen^{104,105}, FVII^{106,107}, FXI¹⁰⁸, TAFI¹⁰⁹, and factor VII-activating protease (FSAP)¹¹⁰.

Epigenetics and hemostasis

A growing body of evidence has indicated an important role for epigenetic mechanisms in the regulation of platelets and plasma proteins involved in blood coagulation.¹¹¹ Over the last few years, technological advances have made it possible to perform so called epigenome-wide association studies (EWASs). An EWAS is performed similarly to a GWAS, but focusing on epigenetic variation (commonly DNA methylation) instead of genetic variation.¹¹² With regards to thrombotic diseases, EWASs have identified several CpGs that are differentially methylated in cases with myocardial infarction (MI) or ischemic stroke compared to controls.¹¹³⁻¹¹⁶ A recent large EWAS also identified differential methylation in relation to cardiovascular risk factors, and a methylation-based risk score was significantly associated with incident cardiovascular events in the Framingham offspring study.¹¹⁷ Another recent EWAS performed on population-based cohorts from the USA and Europe also found that the methylation level of several CpG sites was associated with incident MI and coronary heart disease (CHD), and Mendelian randomization analyses further supported a causal effect of DNA methylation on incident CHD.¹¹⁸ However, in contrast to GWASs on traits such as coronary artery disease (CAD) and stroke that have identified association for variants in several hemostatic genes^{60,119}, EWASs have so far not implicated a role for hemostatic gene DNA methylation in thrombotic disorders.

Role of hemostatic proteins in arterial and venous thrombosis

An increased concentration or activity of circulating prothrombotic proteins, or a decreased concentration or activity of antithrombotic/fibrinolytic proteins, can lead to a prothrombotic state and an increased risk of arterial or venous thrombosis.¹²⁰ Consequently, there have been a large number of studies performed on circulating hemostatic proteins and myocardial infarction, ischemic stroke, and venous thromboembolism (VTE). Associations between increased concentrations of some prothrombotic proteins such as fibrinogen^{121,122}, FVII¹²³⁻¹²⁵ and PAI-1¹²⁶⁻¹³⁰ and incident both myocardial infarction and ischemic stroke have been convincingly demonstrated. Our group and others have also found increased plasma concentrations of several prothrombotic factors in patients that have suffered from ischemic stroke compared to controls.¹³¹⁻¹³⁷ Similar findings have been made for case control studies on CAD.^{138,139}

Role of hemostatic proteins in the central nervous system

Apart from the important functions in the vascular compartment, some hemostatic proteins also influence various processes in the brain. The brain has a unique system for hemostatic regulation, where the antithrombotic and fibrinolytic pathways appear to be less active compared to other tissues, in order to protect against hemorrhage.¹⁴⁰ For example, concentrations of the anticoagulants tissue factor pathway inhibitor (TFPI) and thrombomodulin are very low in the brain^{141,142}, whereas the principal initiator of coagulation, tissue factor, is expressed at high levels by astrocytes^{143,144}.

Among the hemostatic proteins most extensively studied in the central nervous system (CNS) is t-PA. In the CNS, t-PA is not only synthesized in endothelial cells but also in neurons and glial cells such as astrocytes.^{145,146} It has been implicated in several different important physiological processes including synaptic plasticity and memory/learning. t-PA gene expression increases during the late phase of long-term potentiation (L-LTP), a cellular system for memory formation, and in certain neurons in the cerebellum during activity-dependent plasticity.^{147,148} The exact molecular mechanisms by which t-PA facilitates these processes are not yet fully elucidated, but several plausible explanations have been put forward. For example, t-PA is known to associate with certain cell surface receptors including the N-methyl-D-aspartic acid receptor (NMDAR) and the low-density lipoprotein receptor-related protein (LRP). This leads to an enhanced efficiency of intracellular signaling¹⁴⁹ and subsequently in structural remodeling of synapses, synaptic plasticity and activity-dependent learning¹⁵⁰⁻¹⁵². t-PA is also believed to be involved in processes regulating brain plasticity through its ability to cleave and activate plasminogen into plasmin. Active plasmin can, in addition to fibrin degradation, convert the precursor of brain derived neurotrophic factor (BDNF) into mature BDNF, a key protein in the regulation of L-LTP.¹⁵³

Evidence that t-PA is involved in neurotoxic processes has also been proposed. Following cerebral ischemia or traumatic brain injury, large amounts of t-PA are released into the extracellular space by activated glial cells and neurons. This results in cleavage of NMDA receptors and subsequently in an over-excitation of neurons, and neuronal cell death.¹⁵⁴ High concentrations of t-PA may also induce opening of the blood brain barrier (BBB), through a process involving the interaction of t-PA with LRP.¹⁵⁵ Other potential mechanisms have also been put forward, suggesting that t-PA regulates these processes through several different actions.¹⁵⁶ In Paper I, we thus chose to investigate t-PA expression in human neurons and astrocytes.

AIM OF THE THESIS

The overall aim of this thesis was to study the influence of genetic and epigenetic variation on the regulation of hemostatic gene expression.

The specific aims were:

Paper I

To test the hypothesis that epigenetic mechanisms regulate the expression of the gene encoding tissue-type plasminogen activator within the human brain

Papers II and III

To map the gene expression and DNA methylation patterns of hemostatic genes predominantly expressed in the human liver

Papers II and III

To use allele-specific expression and DNA methylation analyses to identify putative *cis*-acting variants involved in hemostatic gene regulation in human liver

Paper IV

To investigate whether DNA methylation patterns in hemostatic genes in blood can reliably predict those in liver

SUBJECTS AND METHODS

Material

The following biological material has been used for the analyses described in this thesis.

Paper I

- Primary human astrocytes derived from two different individuals
- Primary human neurons derived from two different individuals
- Human post-mortem hippocampal and cortical brain tissue from ten individuals

Papers II-IV

- Human liver tissue and peripheral blood samples from 27 adult individuals

Cell culture

For Paper I in this thesis, we utilized primary cultures of two types of human cells derived from the brain. Astrocytes represent a large proportion of all glial cells in the brain.¹⁵⁷ Their main functions include biochemical support of the endothelial cells that form the BBB and to provide neurons with nutrients, but they also play a role in tissue repair following brain injuries.¹⁵⁸ Neurons are the primary cell type in the nervous system where they are specialized in processing and transmission of cellular signals in both chemical and electrical forms. Neurons and astrocytes are both producers of t-PA in the CNS.¹⁴⁵

In vitro culturing of cells is a laboratory technique that is fundamental in many fields of medical research. The cells may be isolated from a living tissue directly, or they may be derived from an already established cell line. A key advantage of *in vitro* experiments is that cells can be maintained and grown under carefully controlled conditions, thereby eliminating many cofounders caused by external factors. However, it should be noted that cultured cells are grown in an artificial setting that cannot accurately simulate the internal environment in the living tissue. Thus, results from cell culture experiments should always be interpreted with caution as they may not be applicable in the *in vivo* context. Furthermore, primary cells have a shorter life span in cultures

compared to established cell lines. However, they are still considered to maintain more of the normal features and functions seen in cells *in vivo*.¹⁵⁹

Cell culture and treatments of human neurons and astrocytes (Paper I)

Human primary astrocytes derived from two individuals (ScienCell, San Diego, CA, USA) were cultured in astrocyte growth medium (AGM) supplemented with 2% fetal bovine serum, 1% astrocyte growth supplement and 1% penicillin/streptomycin solution (ScienCell). Astrocytes were split and subcultured using trypsin-EDTA (0.25 mg/ml, ScienCell) treatment. Human primary neurons (ScienCell) from two individuals were cultured on poly-L-lysine-coated flasks (ScienCell) in neuronal medium supplemented with 1% neuronal growth supplement and 1% penicillin/streptomycin solution (ScienCell). Cultured cells were kept at 37°C and 5% CO₂ in a humidified environment. The medium was replaced in the first two days and every two to three days thereafter.

To evaluate the influence of histone modifications on t-PA gene expression, both types of cells were treated with two different inhibitors of HDAC, trichostatin A (TSA) and MS-275. HDAC inhibitors are a class of compounds that increases acetylation of lysine residues in histones by inhibiting the activity of HDAC enzymes. TSA inhibits class I and class II HDACs, whereas MS-275 specifically inhibits class I HDACs. TSA (1 μM) and MS-275 (10 μM) were prepared in dimethyl sulfoxide (DMSO) and diluted in supplemented astrocyte growth medium for astrocyte cultures or in supplemented neuronal medium for neuron cultures. Control cultures were exposed to the maximum final concentration of DMSO (0.1%). Cells were treated for 14 or 24 hours, and all stimulations were performed in six separate cell culture wells (n = 6) per cell type and treatment time. Astrocyte treatments were performed at passage three and on cells of 100% confluence. Neuron treatments were performed once the seeded cells were fully differentiated. Cell culture media was collected and stored at -20°C. Cells were lysed with TRK buffer (Qiagen, Hilden, Germany) supplemented with 20 μL β-mercaptoethanol per ml TRK, and extracts were then collected and stored at -80°C, until further analysis. Concentrations of TSA and MS-275 were selected to be appropriate for cell treatment with regard to their respective half maximal inhibitory concentration (IC₅₀) measures.

For the chromatin immunoprecipitation (ChIP) assay, primary human astrocytes were cultured as described above, in 15 cm plates and allowed to

reach approximately 80% confluence. Cells were then treated with the same concentrations of TSA (1 μ M), MS-275 (10 μ M), or DMSO vehicle control (0.1%) for 24 hours. Following treatments, the culture medium was removed and replaced by serum-free astrocyte growth medium supplemented with 1% formaldehyde to induce chromatin cross-linking. Plates were incubated with formaldehyde for 10 minutes at room temperature and then washed with ice cold phosphate-buffered saline (PBS). The fixation process was stopped by the addition 1 x glycine/PBS buffer for five minutes and the cells were then washed in cold PBS one more time. Fixed cells were harvested using a rubber policeman and a cell-scraping solution consisting of cold PBS supplemented with 30 μ L of 100 mM phenylmethylsulfonyl fluoride (PMSF), and then pelleted by centrifugation for 10 minutes at 720 RCF and 4°C. The supernatant was discarded, 1 μ L 100 mM PMSF + 1 μ L protein inhibitor cocktail (PIC; Active Motif, Carlsbad, CA, USA) was added, and the pellet was stored in -80°C until further analysis. Since ChIP requires a relatively large amount of cross-linked DNA, this assay could not be performed on neurons, as they do not proliferate in cultures.

Human brain tissue

In addition to the *in vitro* cell culture experiments in Paper I, we also performed DNA methylation analyses on DNA isolated from human hippocampal and cortical brain tissue. Cerebral cortex makes up the outer layer of the brain and is the main site of neural integration.¹⁶⁰ The hippocampus is part of the limbic system and consists of two structures, one in each hemisphere of the brain, and plays a fundamental role in memory consolidation.¹⁶⁰ The benefit of studying whole tissue compared to cultured cells is that it offers a more *in vivo*-like environment, where the biological and chemical nature of the tissue is preserved. However, as whole tissue consists of several different cell types with potentially different DNA methylation signatures, there is always a risk that methylation studies on whole tissue may be confounded by cell-type composition bias.

Human brain tissue collection (Paper I)

Fresh frozen hippocampal and cortical brain tissue samples from 10 unrelated individuals were provided by the Neurological Foundation of New Zealand Human Brain Bank. There were no evident neurological injuries associated with the cause of death, and the assessment of all brains by a neuropathologist

revealed no neurological abnormalities. Mean post-mortem delay was 13 hours (range 7–21 hours).

Human liver

The liver plays a major role in producing proteins that are secreted into the blood, and as outlined in the Background, many circulating proteins involved in hemostasis are primarily produced in the liver. Thus, in Papers II-IV, we studied DNA methylation and gene expression of hemostatic genes in liver.

Human liver tissue and blood sample collection (Papers II-IV)

Human liver tissue and venous blood samples were collected from adult individuals of European descent undergoing liver surgery at the Sahlgrenska University Hospital, Gothenburg, Sweden. The reasons for surgery were secondary malignant neoplasm of the liver and intrahepatic bile duct or liver cell carcinoma. The liver tissue samples collected for this study were as distant as possible to the metastasis/carcinoma. Samples had to be non-cirrhotic and non-tumorous on macroscopical examination by the surgeon to be included in the studies. During the liver resection, non-tumorous tissue (size approximately 0.5 cm³) was collected and placed in RNAlater at 4°C for 3 to 4 days, then aliquoted and stored at –80°C.

Initially, 20 paired liver and blood samples were collected. Genomic DNA (gDNA) isolation and sequencing was successful for 19 blood samples. High quality RNA was successfully isolated and sequenced from 18 liver samples, and both data types were available for 17 paired samples. gDNA extraction from liver tissue and subsequent bisulfite sequencing was successful for all 20 samples. Paper II is based on data from the 17 paired blood DNA/liver RNA samples, and Paper III is based on the 19 paired samples with data from both blood DNA (for genotyping) and liver DNA (for methylation analysis). At the time of the initiation of the study described in Paper IV, we had collected eight additional paired liver and blood samples that were analyzed with regards to genotypes and DNA methylation. Thus, Paper IV included data from 27 paired samples.

mRNA quantification

Quantitative polymerase chain reaction

Quantitative PCR (qPCR) is a sensitive method used for mRNA quantification and expression analysis. Prior to the qPCR, the target mRNA is converted into complementary DNA (cDNA) through a reaction using the DNA polymerase reverse transcriptase. This is performed to achieve a more stable nucleic acid sequence. There are two common versions of qPCR quantification: TaqMan probe-based detection and SYBR green-based detection. In probe-based qPCR, the probe consists of a sequence complementary to a region of the target genes with a reporter fluorophore covalently bound to the 5'-end and a quencher fluorophore bound to the 3'-end. When the fluorophore and quencher are located in proximity to each other, the quencher suppresses the fluorophore signal. The probe is hydrolyzed with the sample, and is allowed to bind to the template strand of the gene of interest. During PCR, the Taq polymerase replicates the template to which the probe is bound, cleaves the probe and thereby separates the quencher from the reporter. The intensity of fluorescence emission increases exponentially as this reaction progresses; this allows for determination of the start concentration of target mRNA on the basis of the emission intensity.

The SYBR green assay uses small molecular fluorogenic probes specific to a target gene for detection as it accumulates during PCR. The probe fluoresces when it is bound to double-stranded DNA. As the DNA is denatured during the PCR, the probe is released and fluorophore emission is reduced to a minimum. During the extension and polymerization process, the fluorescent probe binds to the double-stranded product, resulting in a net increase in the fluorophore signal.

Taq-man qPCR is considered superior to that of SYBR green as it is useful for more applications. Both sensitivity and reproducibility are also considered higher compared to SYBR green. Additionally, as SYBR green binds to double stranded DNA in an unbiased manner, longer PCR products will emit more fluorescence than shorter products. The main disadvantage with TaqMan qPCR is that synthesis of different probes is required for each target, which can be time consuming and less cost-effective in comparison to SYBR green.¹⁶¹

RNA sequencing

mRNA expression can also be calculated from RNA sequencing data. Here, expression is quantified by counting the number of sequence reads mapping to the gene of interest. Counts are then typically normalized by conversion to reads, fragments, or transcripts per kilobase million (RPKM, FPKM, and TPM, respectively). For RPKM, the total number of reads in a sample is counted and this number is then divided by one million (i.e. resulting in a “reads per million” [RPM] factor). The number of reads in a specific gene of interest is then divided by RPM, and thereafter by the length of this gene of interest in kilobases. The last step is performed to adjust for the fact that a longer gene will have more reads than a shorter gene with an identical expression level. The FPKM is identical to RPKM with the exception that FPKM was developed specifically for paired-end sequence data, and thus takes into account that two reads (in a read pair) can correspond to one single fragment. TPM is a newer method, but also very similar to RPKM and FPKM. When calculating TPM, the number of read counts for each gene is divided by its gene length (generating “reads per kilobase”, RPK). After that, the RPK values are summed up within each sample and divided by one million, and the RPK value for the gene of interest is then divided by this factor. This avoids the potential issue with RPKM and FPKM when the sum of normalized reads differs between samples, and facilitates direct comparisons between samples of the proportion of reads mapping to a specific gene of interest.¹⁶² Thus, for our analyses of mRNA expression in Papers II and III, we used the TPM method.

Analysis of t-PA mRNA expression using qPCR (Papers I and II)

In Paper I, TaqMan Real-Time qPCR (Applied Biosystems, Foster City, CA, USA) was used for mRNA expression analysis of t-PA. Total cell RNA was isolated from cell extracts using E.Z.N.A. Total RNA Isolation Kit (Omega Bio-tek, Norcross, GA, USA), according to the manufacturer’s protocol. The concentrations of the isolated RNA samples were measured by NanoDrop and determined to be approximately 100 ng/μL per sample. Isolated RNA was converted to cDNA using a GeneAmp RNA PCR Kit (Applied Biosystems) in a total volume of 20 μL. All samples were normalized relative to the housekeeping gene glyceraldehyde 3-phosphate dehydrogenase. The PCR was performed in 384-well format on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems) in a total volume of 10 μL. Relative quantification of mRNA expression was analyzed using the comparative CT method and all samples were analyzed in triplicates. Thermocycling conditions and probe and primer sequences are thoroughly described in Paper I.

Analysis of hemostatic mRNA expression (Papers II and III)

In Paper II and III, RNA-seq was used to determine mRNA expression of the targeted genes of interest. Total RNA was extracted from liver tissue samples in biological duplicates (two separate pieces of liver tissue per individual) using the AllPrep DNA/RNAmini kit (Qiagen) and sequenced as described below. Sequence reads were counted using HTseq (v. 0.6.1) in R, and normalized estimates of gene expression in TPM per gene were then calculated.

Protein analyses

There is a variety of different methods for protein quantification. For the protein measurements in Paper I, we utilized a commonly used antibody-based method called enzyme-linked immunosorbent assay (ELISA). A microtiter plate is coated with a known quantity of a capture antibody with specificity for the sample antigen. Samples are then added to the wells and are bound by the capture antibody. The plate is then washed to remove any unbound antigen before a secondary enzyme-linked antibody specific for the antigen is added. Finally, a substance with the substrate for the enzyme is added and the reaction generates a detectable signal that will be based on the amount of bound secondary antibody and thereby the amount of antigen.

Analysis of t-PA protein concentrations (Paper I)

Cell culture medium from astrocytes treated with HDAC inhibitors and control cultures was collected and t-PA protein concentrations in the medium was determined using a sandwich ELISA (TriniLIZE t-PA; Trinity Biotech, Bray, Ireland) according to the manufacturer's protocol. The colour intensities of the samples were measured at 490 nm in a Mithras LB 940 (Berthold Technologies, Bad Wildbad, Germany).

Chromatin immunoprecipitation assay

Chromatin immunoprecipitation (ChIP) assay is a method used to map the DNA target of specific DNA-binding proteins, such as transcription binding factors and histones. The technique is based on cross-linkage of the entire chromatin in a tissue or living cells by fixation. The cross-linked chromatin is then sheared by enzymes or sonication into smaller fragments, and the regions bound to the protein of interest can be extracted using a protein-specific antibody. Cross-linkage is then reversed and the DNA is released, purified, and analyzed.

Analysis of histone modifications in the t-PA promoter (Paper I)

Cell pellets were thawed on ice and resuspended in cold lysis buffer supplemented with PMSF and PIC and incubated on ice for 30 minutes. Cells were transferred to a cold homogenizer and dounced on ice with a small clearance pestle and 10 strokes to aid nuclei release. The samples were then centrifuged for 10 minutes at 2,400 RCF at 4°C and the supernatant was discarded. The nuclei pellet was resuspended in shearing buffer supplemented with PMSF and PIC, and sheared on a Bioruptor Pico (Diagenode, Liège, Belgium) sonicator with 25% power at 4°C, with five pulses of 20 seconds each, with 30 seconds rest in between. Samples were then centrifuged for 10 minutes at 18,000 RCF and 4°C, and the supernatant containing the sheared chromatin was collected. Ten µL of each sample was removed and stored at -20°C to serve as the input control.

The ChIP reaction was performed with rabbit anti-acetyl-histone H3 or anti-acetyl-histone H4 (1 mg/mL; Merck Millipore, Burlington, MA, USA). A negative antibody control (normal rabbit IgG), and a no-antibody control were also prepared. All samples were incubated overnight on an end-to-end rotator at 4°C, and the captured DNA was eluted by magnetic separation of the beads.

Both ChIP samples and input DNA samples were incubated at 95°C for 15 minutes. Following incubation, proteinase K was added to each tube, and samples were incubated at 37°C for one hour. Thereafter, the tubes were returned to room temperature and proteinase K stop solution was added to each sample. Finally, samples were cleaned using the Chromatin IP DNA Purification Kit (Active Motif) as recommended by the manufacturer, and the captured DNA was quantified using SYBR green detection. Samples were analyzed in triplicates on a CFX384 Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA). A standard curve with 7 points

(concentration ranging from 3.2 pg/μl to 50 ng/μl) and a no-templet control for the qPCR were included in triplicates. The results were calculated as the percentage of input fraction, corrected for background with no-antibody controls. The SYBR green qPCR concept is described in detail above. The qPCR was performed at TATAA Biocenter in Gothenburg, Sweden. Primer sequences and further details can be found in Paper I.

Methylation sequencing of the t-PA promoter (Paper I)

Total DNA from cultured human astrocytes, neurons, and hippocampal and cortical brain tissues was isolated using the E.Z.N.A. Tissue DNA Kit (Omega Bio-tek). Isolated DNA was modified with sodium bisulfite using a MethylCode Bisulfite Conversion Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's recommendations. Following bisulfite conversion, three overlapping regions of the t-PA promoter were amplified by a so called "nested PCR", in which the DNA is amplified in two subsequent PCRs. The first PCR is performed with a pair of outer primers flanking the target sequence, followed by a second PCR utilizing a pair of inner primers. This method has the advantage that it reduces the risk of contamination due to unexpected primer binding sites, and provides a higher overall yield than a traditional PCR. Conditions for both the primary and the secondary reaction as well as primer sequences are reported in detail in Paper I. After each stage of the nested PCR the product was purified using the High Pure PCR Product Purification kit (Roche, Basel, Swizerland). The PCR product was then sequenced directly by Sanger sequencing on an ABI 3130 XL (Applied Biosystems), using the forward primers from the nested PCR. Percent methylation of all CpGs in the promoter was analyzed using the 4Peaks (Nucleobytes, Aalsmeer, The Netherlands) software.

Design of a targeted sequencing panel for hemostatic genes (Papers II-IV)

For Papers II-IV, we selected 35 genes with predominant expression in liver based on data from the GTEx project⁶⁶, and related to hemostasis by studies of the literature, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database¹⁶³, and the Molecular Signatures Database (MSigDB)¹⁶⁴. To exploit the full potential of the design, we also included 42 other genes of interest for our group. Custom capture kits (Agilent, Santa Clara, California, United States) were designed for targeted sequencing of gDNA and for methylation analysis using the SureDesign tool (Agilent), and for mRNA using eArray (Agilent). For gDNA methylation sequencing, probes were designed to capture all exons, introns, and potential upstream and downstream regulatory regions, resulting in a 3.3 Mb design. Genomic regions 5,000 bp upstream and 500 bp downstream of each gene were included in the design. However, for the upstream regions we expanded the default 5,000 bp region if we found strong evidence for more distal regulatory elements further upstream. For the RNA design, exons from all known isoforms of the 35 genes were included, plus one negative control and four liver cell-specific markers, resulting in a 0.5 Mb design. A full list of genes, regions, and transcripts for the 35 genes included in the gDNA and methylation sequencing designs, and the negative control and liver-specific markers included in the RNA design are shown in **Table 1**.

Table 1. Genes and targeted regions included in the custom-designed capture kits for gDNA, mRNA and methylation sequencing. Locations are given according to the human reference genome hg19.

Gene name	Region included in the capture design	Protein name
<i>A2M</i>	chr12:9,219,804-9,280,558	Alpha2-macroglobulin
<i>C4BPB</i>	chr1:207,257,584-207,273,837	Complement component 4 binding protein, beta
<i>CPB2</i>	chr13:46,626,822-46,684,211	Carboxypeptidase B2 (TAFI)
<i>F10</i>	chr13:113,774,613-113,804,343	Factor X
<i>F11</i>	chr4:187,180,618-187,211,335	Factor XI
<i>F12</i>	chr5:176,828,639-176,844,077	Factor XII
<i>F13B</i>	chr1:197,007,821-197,041,397	Factor XIII B
<i>F2</i>	chr11:46,732,043-46,761,556	Prothrombin (Factor II)
<i>F5</i>	chr1:169,480,692-169,560,769	Factor V
<i>F7</i>	chr13:113,755,105-113,775,495	Factor VII
<i>F9</i>	chrX:138,607,895-138,646,117	Factor IX
<i>FGA</i>	chr4:155,503,780-155,516,897	Fibrinogen alpha chain
<i>FGB</i>	chr4:155,479,132-155,494,415	Fibrinogen beta chain
<i>FGG</i>	chr4:155,525,228-155,539,402	Fibrinogen gamma chain
<i>GGCX</i>	chr2:85,771,478-85,798,657	Vitamin K-dependent gamma-carboxylase
<i>HABP2</i>	chr10:115,305,590-115,349,860	Factor VII activating protease (FSAP)
<i>HRG</i>	chr3:186,378,798-186,396,523	Histidine-rich glycoprotein
<i>KLKB1</i>	chr4:187,143,672-187,180,125	Plasma kallikrein
<i>KNG1</i>	chr3:186,415,098-186,461,178	Kininogen-1
<i>LMAN1</i>	chr18:56,994,556-57,031,508	Lectin, mannose-binding, 1
<i>MCFD2</i>	chr2:47,128,509-47,154,215	Multiple coagulation factor deficiency protein 2
<i>PLG</i>	chr6:161,118,225-161,175,585	Plasminogen
<i>PROC</i>	chr2:128,142,996-128,187,322	Vitamin K-dependent protein C
<i>PROS1</i>	chr3:93,591,381-93,697,934	Vitamin K-dependent protein S
<i>PROZ</i>	chr13:113,805,468-113,827,194	Vitamin K-dependent protein Z
<i>SERPINA1</i>	chr14:94,842,584-94,862,029	Alpha1-antitrypsin
<i>SERPINA10</i>	chr14:94,749,150-94,764,608	Protein Z-related protease inhibitor
<i>SERPINA5</i>	chr14:95,041,731-95,059,957	Plasma serine protease inhibitor
<i>SERPINC1</i>	chr1:173,872,442-173,891,516	Antithrombin III
<i>SERPIND1</i>	chr22:21,123,383-21,142,508	Heparin cofactor 2
<i>SERPINE1</i>	chr7:100,764,879-100,783,047	Plasminogen activator inhibitor-1
<i>SERPINF2</i>	chr17:1,636,130-1,659,059	Alpha2-antiplasmin
<i>SERPING1</i>	chr11:57,360,027-57,382,826	Plasma protease C1 inhibitor
<i>TFPI</i>	chr2:188,328,458-188,424,219	Tissue factor pathway inhibitor
<i>VKORC1</i>	chr16:31,101,675-31,111,276	Vitamin K epoxide reductase complex subunit 1
Liver-specific markers included in the RNA design		
<i>ALB</i>		Serum Albumin
<i>CD163</i>		CD163 molecule
<i>FBLN2</i>		Fibulin-2
<i>CD3G</i>		CD3 gamma molecule
Negative control included in the RNA design		
<i>HYAL4</i>		Hyaluronidase 4

Massive parallel sequencing

The workflow for preparing custom sequencing libraries for massive parallel sequencing is illustrated in **Figure 4**.

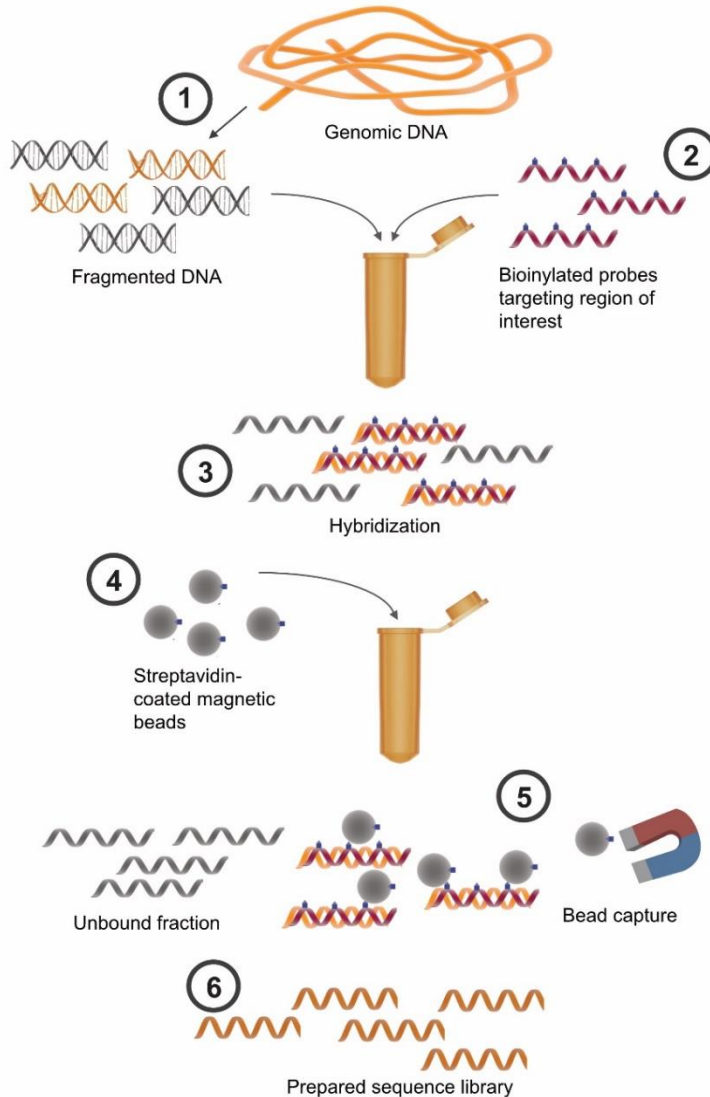


Figure 4. Library preparation for targeted next generation sequencing. 1) Genomic DNA is sheared into smaller fragments. 2) Biotinylated probes are designed to target a specific region of interest. 3) Probes are hybridized with the genomic sample and will only anneal to the targeted regions. 4) Streptavidin-coated magnetic beads are added to the sample. 5) The Streptavidin-coated beads bind to the biotinylated probes and the target DNA is extracted through magnetic separation. 6) Targeted DNA is eluted and the sequence library is ready.

DNA sequencing of hemostatic genes (Papers II-III)

gDNA from whole blood was isolated using the QIAamp DNA Blood Midi Kit (Qiagen). Strand-specific sequencing libraries for gDNA-seq were prepared using a custom designed SureSelect kit (Agilent) according to the manufacturer's protocols. Input amount of gDNA was 3 µg. Six or seven samples were pooled into one single 8 pM sequence library and multiplexed on one flowcell on the MiSeq platform (Illumina, San Diego, CA, USA) with 150 bp paired-end reads.

All raw sequence reads were quality controlled using FastQC (Babraham Bioinformatics, Cambridge, UK). Reads were trimmed (removal of adaptors and indexes) with Prinseq-lite (v0.20.3) and aligned to the human reference genome hg19 with Bowtie2 (v2.2.6). Genetic variants were called using HaplotypeCaller v3.4.36 (GATK, Broad Inst, Boston, Massachusetts, USA) with the GATK Best Practice recommendations, using a minimum required quality score of ≥ 40 .

RNA sequencing

Because of the sensitive nature, RNA needs to be converted to cDNA before sequencing. This is performed with a reverse transcriptase in a mix with the four nucleotides in a PCR. During the reaction, a complementary strand of DNA is synthesized using the mRNA as a template.

RNA sequencing of hemostatic genes (Papers II-III)

Total RNA from liver tissue samples were extracted in biological duplicates (i.e., two separate pieces of liver tissue per individual) using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). Strand-specific RNA-seq libraries were prepared for each duplicate separately using the custom-designed SureSelect kit (Agilent) according to the manufacturer's protocol. A total of 10-12 sample libraries were multiplexed per sequence lane using 125 bp paired-end reads on the HiSeq 2500 sequencer (Illumina). Samples from two subjects 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).

Sequence reads were trimmed with TrimGalore! (v.0.4.1) and aligned via STAR (v2.4.2) to N-masked individual-specific versions of the reference genome hg19, constructed using genotype data from the same samples. This

was performed to minimize downstream mapping biases for the ASE analyses, as recommended by GATK ASE best practices. Reads were counted for mRNA quantification using HTSeq (v.0.6.1) using default parameters and the results were compared between duplicates and between the custom and commercial library kits.

Methylation sequencing

Sodium bisulfite treatment of DNA converts all unmethylated cytosines in the sequence to uracils. This is achieved when bisulfite induces deamination of unmethylated cytosines in single stranded DNA, while methylated cytosines are protected against deamination. By sequencing the bisulfite treated gene and comparing it to an untreated reference sequence, it is possible to determine the DNA methylation status of the methylation sites in question. An illustration of bisulfite conversion is shown in **Figure 5**.

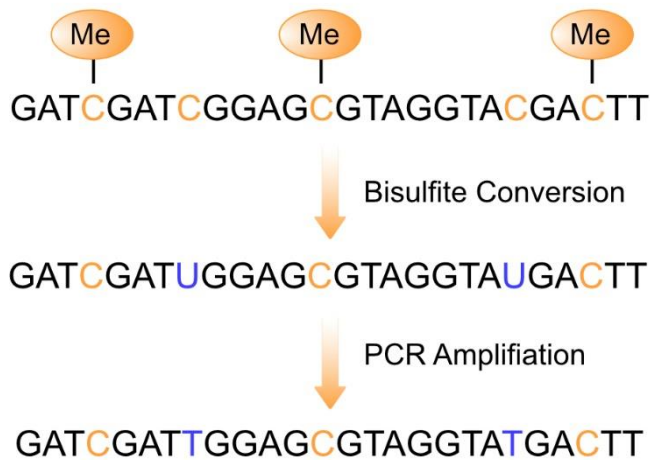


Figure 5. Bisulfite conversion. Treatment with sodium bisulfite converts all unmethylated cytosines (C) to uracils (U), whereas the methylated cytosines remain unchanged. During the PCR, thymine (T) will be incorporated instead of the uracils, and the methylation status of each CpG can be determined.

Methylation sequencing of hemostatic genes (Papers III-IV)

gDNA from blood was isolated as described above, and gDNA from liver tissue was isolated using the AllPrep DNA/RNAmini kit (Qiagen). The DNA was sheared to an average size of 250 bp and bisulfite converted with the EZ DNA methylation kit (Zymo Research, Irvine, CA, USA). 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 HiSeq 2500 (Illumina) with High Output mode and 100 bp paired-end reads, following the manufacturer's protocols. For comparison, gDNA isolated from two liver samples was also prepared using the SureSelectXT Methyl-seq target enrichment system for Illumina (Agilent), which is a commercially available methylation sequencing kit targeting 84 Mb of the genome, and sequenced in one lane on the HiSeq 2500 with the same options as described above.

All raw output sequence reads were quality controlled using FastQC (Babraham). After removal of low-quality bases (Phred score <30), reads were trimmed with TrimGalore! (v.0.4.1) and aligned to an *in-silico* bisulfite converted version of the human reference genome hg19 with Bowtie2 (v. 2.2.6.) Deduplication and methylation calling were performed with the Bismark pipeline (v.0.14.5) using default parameters. The SeqMonk software (v1.44.0; Babraham) was applied to visualize CpG methylation levels.

Sequence data quantitation and downstream analyses

Sequencing libraries for gDNA were prepared and sequenced at the Genomics Core Facility at the University of Gothenburg. For mRNA, a pilot of six samples were initially prepared and sequenced at the Genomics Core Facility. After evaluating the data, we prepared libraries for all samples (in biological duplicates), which were then sequenced at the National Genomics Infrastructure (NGI) at Science for Life Laboratory. Methylation sequencing libraries were both prepared and sequenced at NGI.

Allele-specific expression analysis

As conventional eQTL studies require large sample sizes, only few studies have been performed on human liver tissue. Furthermore, results from eQTL analyses are influenced by unmeasured confounding variables, and thus the reproducibility of eQTL studies is relatively low.^{165,166} The ASE approach is an alternative methodology to map putative SNPs affecting mRNA expression.

This method standardizes environmental and other confounding variables and is a powerful method for small sample sizes which facilitates analysis of more “hard to access” human tissues.⁷² For the purpose of studying genotype-expression association in human liver tissue in Paper II, we thus chose the ASE approach.

Analysis of allele-specific expression of hemostatic genes (Paper II)

Read counts from the biological duplicates were merged prior to ASE analysis. ASE was then determined based on the allelic ratio of heterozygous SNPs in processed RNA sequencing data. The GATK HaplotypeCaller (v.3.4.36) was used to determine RNA zygosity using a minimum required quality score of ≥ 60 . The proportion of reference alleles versus total counts was calculated using ASEReadCounter v.3.4.46 (GATK), following GATK best practices where applicable. For the ASEReadCounter, a read depth of ≥ 10 and a quality score of ≥ 10 was required.

Allele-specific DNA methylation analysis

Since DNA methylation may be influenced by environmental factors (e.g. smoking, diet, and drugs)¹⁶⁷, large sample sizes are required in order to study variation in methylation in relation to genetic variants between individuals. 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. For the same reason, most mQTL studies have used high-throughput assays based on microarrays that cover less than 4% of all human CpG and mainly assay CpGs in promoters and CGIs.¹⁶⁸ The alternative 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. For these reasons, we chose the ASM approach for the genotype-methylation studies in Paper III.

Analysis of allele-specific DNA methylation in hemostatic genes

Bisulfite-seq reads were aligned to an *in-silico* bisulfite converted version of hg19 reference genome, in which known SNP positions had been masked by the ambiguity base “N” prior to alignment to avoid potential read mapping biases toward higher mapping rates of the reference allele compared with the

alternative allele¹⁶⁹, using Bowtie2 (v.2.2.6). SNPsplit (Babraham) was then used to determine the allelic origin of reads covering known SNP positions using individual-specific genotype data. Aligned reads were separated into two separate alignments (i.e. one major allele alignment and one minor allele alignment) based on the alleles of the overlapping SNPs. The methylation status was then called from the two alignments separately using Bismark (v.0.14.5), and the methylation level of each CpG was compared between the two alignments. In order to appropriately estimate methylation differences, we required a minimum of 30 reads per allele.

Databases and tools

The sequencing projects described earlier, and many others, make their data available for the research community in various ways. A considerable number of databases and browsers have been constructed to compile data from different studies and projects. For the purpose of this thesis, I will mention a subset of these:

The Single Nucleotide Polymorphism Database (dbSNP)¹⁷⁰: Contains information on genetic variation for several species, including humans, and is developed and hosted by the National Center for Biotechnology Information (NCBI) in collaboration with the National Human Genome Research Institute (NHGRI). This database was used for SNP annotation in Papers II and III.

The Genotype-Tissue Expression (GTEx) project⁶⁶: A database of correlations between genotype and tissue-specific gene expression that is based on RNA sequencing data from several human tissues from genotyped donors. GTEx was used to assess liver expression of the targeted hemostatic genes investigated in Papers II-IV

Ensembl¹⁷¹: A genome browser with reference datasets, with much focus on annotation. Ensembl is part of the European Bioinformatics Institute (EMBL-EBI) at the European Molecular Biology Laboratory (EMBL). Ensembl was used for exon annotation in Paper II, and for genomic context annotations in Papers II-IV.

The NHGRI GWAS catalog¹⁷²: A searchable catalog of published genome-wide association studies, founded by NHGRI and hosted by the EMBL-EBI. This library was used to assess if any of the expression and methylation associated SNPs identified in Paper II and III were associated with circulating levels of the respective proteins or other relevant traits.

Phenoscaner¹⁷³: A database of results from published human GWAS, eQTL, pQTL, mQTL, and metabolite QTL data. This database was used to assess if the identified ASE and ASM-SNPs in Papers II and III had previously been associated with circulating hemostatic factors or other relevant traits.

University of California, Santa Cruz (UCSC) genome browser¹⁷⁴: An interactive website with genome sequence data integrated with a large collection of annotations from various projects. The UCSC genome browser was used to assay potential regulatory regions around hemostatic genes in order to determine the upstream and downstream distance to be included in the targeted sequencing design used in Papers II-IV.

HaploReg¹⁷⁵: A tool based on data from ENCODE for the chromatin state and protein binding annotation of SNPs. We used this tool for assessments of promoters and enhancers overlapping the ASE- and ASM-SNPs identified in Papers II and III.

Statistical analyses

In Paper I, differences in mRNA levels, protein concentrations, and degree of histone acetylation between HDAC inhibitor-treated cells and control-treated cells were determined by the unpaired Student's t-test. A *P*-value <0.05 was considered statistically significant. Statistical analyses were performed using IBM SPSS for Windows (v.20, release 20.0.0).

In Paper II, allele-specific expression was estimated by binomial testing in ASEReadCounter (v3.4.46; GATK). Further, for variants to be classified as ASE sites, the allelic imbalance ratio had to be <0.35 or >0.65 in at least one individual, with an average of ≤ 0.4 or ≥ 0.6 between all heterozygous individuals.

In Paper III, differences in methylation levels between alleles for each individual were weighted by their inverse variances and summed. The sum was then scaled by its standard error to form a Z-statistic. A *P*-value less than 0.05 was considered statistically significant. Further, a percent methylation difference >20% between the two alleles was required in order for CpGs to be classified as displaying ASM. Statistical analyses were performed using R (v. 3.6.1).

In Paper IV, we used Pearson's correlation to evaluate correlations in methylation levels between paired liver and blood samples. A minimum of five

paired liver-blood samples per CpG was set as a requirement for this test and a Bonferroni adjusted P -value of <0.05 was considered statistically significant. In addition to Bonferroni correction for multiple testing, we also performed a less conservative control of Type I errors using false discovery rate (FDR), with a significance cutoff of 0.05. Correlations for average methylation across all subjects between liver and blood were calculated using Pearson's correlation. All statistical analyses in Paper IV were performed using R (v. 3.6.1).

Ethical approvals and considerations

For Paper I, all donated brain tissues were collected with the full consent of the next of kin and in line with ethics approval from the University of Auckland Human Participants Ethics Committee (ref 011654). For Papers II-IV, written informed consent was obtained from all participants, and the study was approved by the Ethics Committee at the University of Gothenburg (dnr 665-13).

There were no medical risks for patients participating in the studies. Thus, the only relevant risks concern patient integrity with regard to personal and genetic information. To overcome this, all samples were coded immediately after collection and cannot be directly traced to the participant's name or social security number. The genetic analyses did not provide participants with any information of relevance for their prognosis or for risk factors regarding other diseases. There are no foreseeable benefits for the patients that chose to participate. However, knowledge obtained in these studies may be of value for other patient categories further on.

RESULTS AND DISCUSSION

Regulation of t-PA in the human brain and brain-derived cells (Paper I)

Following treatment with the HDAC inhibitors TSA and MS-275 for 24 hours, we observed a 2 to 3-fold increase in t-PA mRNA in astrocyte and neuron cultures compared to the control-treated cells. At the protein level, we also found that t-PA antigen concentrations in the astrocyte-conditioned media increased 1.5 to 2-fold in response to HDAC inhibitor treatment (**Figure 6**).

As HDACs also act on non-histone proteins that may influence gene regulation, such as transcription factors, we next performed a ChIP assay on astrocytes with antibodies directed against acetylated histones H3 and H4 (both markers of gene activation). Treatment with MS-275 and TSA resulted in a significant increase in H3 acetylation in the t-PA promoter as compared to control treated cells (**Figure 7**). This finding confirms that HDAC inhibitors act directly on histone H3 acetylation in the t-PA promoter, and this mechanism could thus explain the observed increase in t-PA gene expression after the inhibition of histone deacetylation. No significant increase in histone H4 acetylation was observed following HDAC inhibitor treatments.

In cortical and hippocampal brain tissues, as well as in cultured astrocytes and neurons, we observed a high degree of DNA methylation (50–100%) in the t-PA promoter from position –618 to –366 relative to the transcription start site (TSS), whereas the CpGs located between position –121 to +94 showed no or minimal methylation (0–20%). The observed hypomethylation at the latter positions is in line with previous results from human endothelial cells, whereas human hepatocytes exhibit a considerably higher degree of DNA methylation in the same genetic region.¹⁷⁶ Neurons, astrocytes, and endothelial cells have all been shown to have a high basal t-PA gene expression^{145,177}, whereas hepatoma cells and hepatocytes have a low t-PA gene expression¹⁷⁶. Thus, the expression level of t-PA appears inversely correlated with the degree of promoter methylation, which is expected for genes that are under epigenetic influence. Taken together, our results suggest that epigenetic mechanisms are involved in the regulation of t-PA in the human brain. However, the significance of these mechanisms needs to be examined further.

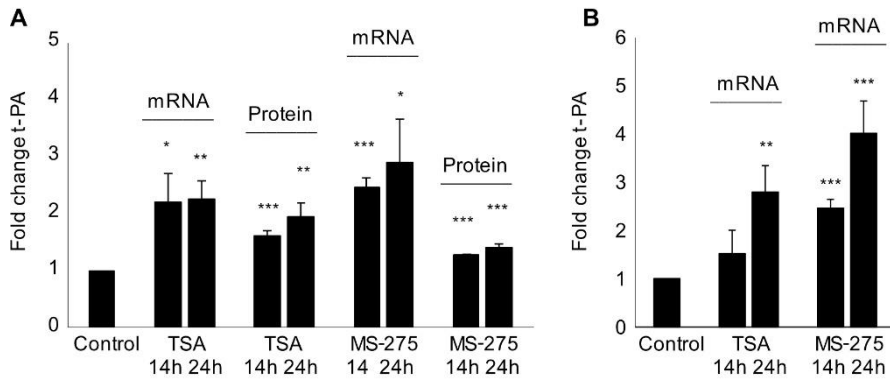


Figure 6. Treatment with histone deacetylase (HDAC) inhibitors resulted in a significant increase in t-PA expression in cultured human astrocytes and neurons. **(A)** t-PA mRNA and protein expression levels in cultured human astrocytes following treatment with 1 μ M trichostatin A (TSA) or 10 μ M MS-275. **(B)** t-PA mRNA expression levels in cultured human neurons following treatment with TSA or MS-275. Cells were treated for 14 or 24 hours. Results are shown as fold induction compared with control-treated cells and presented as mean \pm SEM. Each data point represents the average of three independent treatment series performed on two different occasions. Response to treatment was evaluated by the unpaired Student's t-test: * $P \leq 0.05$, ** $P \leq 0.01$, and *** $P \leq 0.001$.

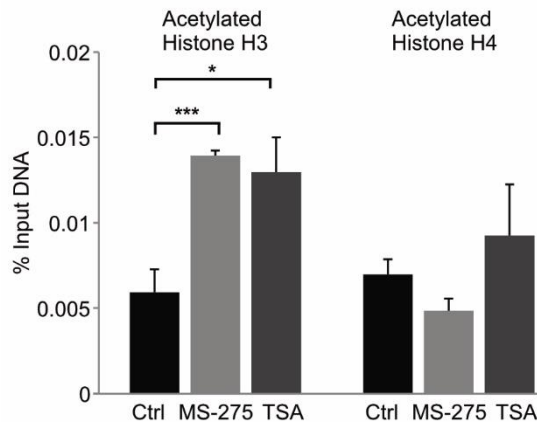


Figure 7. Treatment with histone deacetylase (HDAC) inhibitors resulted in a significant increase in acetylation of histone H3 but not histone H4 in the t-PA promoter. Results are presented as mean percentage of input DNA \pm SEM. Each data point represents the average of two independent ChIP experiments performed in triplicate. Differences in histone acetylation levels were determined by the unpaired Student's t-test. * $P \leq 0.05$, ** $P \leq 0.01$, and *** $P \leq 0.001$.

Sequencing output and evaluation of the targeted design (Papers II-IV)

We made a custom design for the targeted analyses of 78 genes expressed in human liver, and 35 of these were selected for the analyses of hemostatic genes predominantly expressed in liver in Papers II-IV. This design for DNA sequencing yielded approximately 300 million (M) reads per sample, with a sequencing depth of ~100x.

For mRNA sequencing, approximately 25 M reads were generated for each sample. All transcripts of interest were identified. The negative control was not detected. All RNA samples were sequenced in biological duplicates. Read counts and normalized expression values were highly replicable between these duplicates (**Figure 8A**). Two samples were also sequenced using Illumina's TruSeq kit to identify possible weaknesses or biases in our custom designed kit. Assessment of the data showed no significant bias in expression values between the two kits (**Figure 8B**). Furthermore, the custom design gave more sequencing coverage on our transcripts of interest, and thus this approach was more data efficient and cost-effective.

For methylation sequencing, approximately 10-20 M reads per sample were generated, and the sequencing depth was 50x. We evaluated the sequencing output when using the commercial 84 Mb kit compared to the custom kit. The 84 Mb kit is biased towards CpG islands, promoters and regulatory regions. However, according to the manufacturer's specifications, all of our regions of interest (i.e. the 35 hemostatic genes) are included in this design. We compared the depth and coverage within our targeted regions between the two kits. The comparison showed that our custom kit was superior to the commercial kit in both aspects (**Figure 9**).

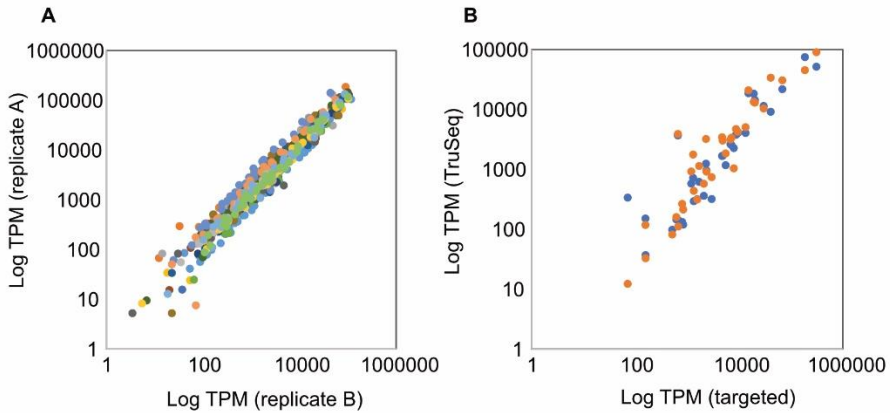


Figure 8. RNA expression data. **(A)** Comparison of normalized expression values (TPM) for each of the 35 genes between duplicates for all 17 samples. **(B)** Comparison of expression values of the 35 genes for one sample prepared and sequenced using our custom capture kit and the TruSeq whole transcriptome kit (Illumina).

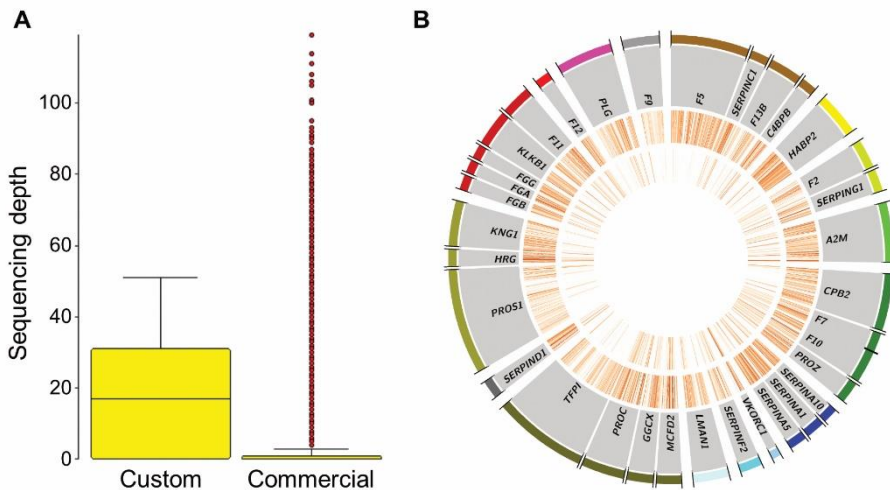


Figure 9. Comparison of sequencing depth and coverage for data generated with the custom-designed capture kit for methylation sequencing and the commercial SureSelect enrichment kit (Agilent). **(A)** Sequencing depth over CpGs within the targeted 35 regions. **(B)** Sequencing coverage for one sample prepared and sequenced using our custom capture kit (outer data track) and the commercial kit targeting 84 Mb genome-wide (inner data track). Sequencing depth throughout the 35 targeted regions is represented by the data track colour intensity, where darker red indicates greater sequence coverage.

Hemostatic genes exhibit allele-specific expression in human liver tissue (Paper II)

Using the targeted DNA- and RNA-sequencing described above, our study identified ASE in 21 of the 35 analyzed hemostatic genes, across a total of 53 SNPs, in liver. Using publicly available database sources¹⁷³, we found that seven of these ASE-SNPs had previously been associated with expression of the respective mRNA in liver tissue in eQTLs studies, and 32 had been identified as eQTLs in other cell types or tissues. As the sample size greatly affects identification of eQTLs, more readily available tissues, such as blood, tend to be overrepresented in eQTL datasets. This is likely the reason for the fact that a much higher number of ASE-SNPs had previously been identified as eQTLs in other tissues, as compared to liver.

More importantly, we also identified novel genotype–expression associations. Fourteen (26%) of the ASE-SNPs identified in this study had no previous evidence of association to mRNA expression, either in the liver or in other tissues. We assessed the novel ASE-SNPs for overlap with regulatory motifs and chromatin states in their genomic location in liver tissue. According to HaploReg¹⁷⁵, 13 had chromatin states indicative of promoters or enhancer elements in the liver, and 11 of these were predicted to abolish, create, or change the affinity of one or several transcription factor binding sites. Five of the novel ASE-SNPs had previously been associated with circulating levels or activity of their respective proteins in publicly available GWAS data¹⁷³, and two additional ASE-SNPs could be associated with concentrations of their respective protein in another published study¹⁰⁹ or in data from our own case–control study on ischemic stroke, SAHLIS¹⁷⁸. Furthermore, seven novel ASE-SNPs, including two of the ones associated with circulating proteins, had previously been associated with other traits or diseases related to thrombosis and hemostasis, including activated partial thromboplastin time, bleeding, CAD, MI, and VTE in public GWAS data.¹⁷³ Taken together, our assessment of previously published data provided evidence for involvement in thrombotic or hemostatic processes for 11 of the 14 novel ASE-SNPs. The remaining three SNPs have no previous mention in the literature, as far as we can tell. These were located in the genes encoding coagulation FX (*F10*; rs3212994), kininogen 1 (*KNG1*, rs5030100), and alfa 1-antitrypsin (*SERPINA10*; rs201774333). However, the *F10* SNP had a chromatin state indicative of a promoter in liver tissue and was predicted to alter two regulatory motifs according to HaploReg. The ASE-SNP in *KNG1* had an enhancer chromatin signature in liver, and was predicted to alter the binding site of three transcription factors. The *SERPINA10* SNP had an enhancer chromatin

signature. Thus, all novel ASE-SNPs that we identified have features indicative of a plausible biologically relevant function and may be candidates for future functional studies. We envisage that this data set could be used by the hemostatic research community. For example, if a novel association between a SNP in one of the 35 hemostatic genes investigated here and a thrombotic or bleeding trait is discovered, the present data can be used to evaluate if this SNP displays ASE in the liver. In this way, it could be a stepping stone for designing appropriate experimental studies. Our results further provide support for the strengths of the allele-specific approach in studies where collection of the relevant tissue, and thereby sample size, is a limiting factor.

Allele-specific DNA methylation of hemostatic genes in human liver (Paper III)

As a next step, we investigated the presence of ASM in liver for the same 35 hemostatic genes. We identified 116 CpG sites displaying significant allelic methylation imbalance of 20% or more in at least one sample. These CpGs were found in 24 genes, and were associated with 112 SNPs. The majority were located in gene bodies, and an enrichment of ASM was observed in the genomic context “first exons”. This is in line with previous studies showing that CpG sites that are associated with expression levels are enriched in enhancers, gene bodies, CpG island shores, and in the first exon, but not in promoter regions.^{16,179} Nearly all (94%) of the ASM-associated SNPs (ASM-SNPs) in our study had chromatin states indicative of promoter or enhancer elements in liver, which is also in line with previous studies in other tissues that have shown that *cis*-acting methylation-associated SNPs are enriched in regulatory chromatin domains and transcription factor binding sites.¹⁸⁰⁻¹⁸²

We investigated whether the identified ASM-SNPs and high LD proxies ($r^2 \geq 0.8$ in 1000G European panel) were associated with mRNA expression using mRNA-seq data derived from the same individuals. However, as we were underpowered for inter-individual analysis, we also queried these ASM-SNPs in a publicly available database¹⁷³ of liver tissue eQTLs. We found evidence of association with mRNA expression in liver tissue for 22% of the identified ASM-SNPs using these approaches. Furthermore, 46% ASM-SNPs had been identified as eQTLs for their respective genes in tissues other than liver. Thus, for 68% of the identified ASM-CpG, our findings suggest that DNA methylation regulated by genetic variants in *cis* may be a mechanistic link between previously identified genotype-expression associations.

We further investigated if the ASM-SNPs had been associated with circulating plasma proteins. Twenty-nine percent had been associated with the respective protein concentrations or activity in publicly available pQTL and GWAS datasets¹⁷³, and these included five ASM-SNPs with no previously reported eQTLs that were associated with plasma fibrinogen, KNG1, SERPINF2, and FSAP concentrations. We also assessed associations of ASM-SNPs with relevant disease and non-disease phenotypes in published data. Two ASM-SNPs in *F5* (rs3766110 and rs3766111) had been associated with VTE (<http://www.ukbiobank.ac.uk>), one ASM-SNP in *TFPI* (rs8176546) had been associated with CAD¹⁸³, and one in *F2* (rs2070851) had been associated with ischemic stroke¹⁸⁴. Furthermore, 34 additional ASM-SNPs had been associated with other phenotypes related to hemostasis such as plasma concentrations of glycoproteins and various coagulation factors, and vascular or heart problems.

The CpG displaying the highest degree of ASM in our study was located in *VKORC1*, with a methylation difference of 47% in favor of the minor allele. The SNP (rs8050894) connected to this CpG was associated with *VKORC1* mRNA levels both in our RNA-seq data and in public liver eQTL data, and it was also the top hit in a GWAS on warfarin maintenance dose¹⁸⁵. Several clinical trials are investigating the value of genotype-guided dosing of oral anticoagulants such as warfarin and novel oral anticoagulants (NOACs)¹⁸⁶⁻¹⁸⁹, and another SNP in almost perfect LD with the *VKORC1* ASM-SNP has already been used in trials of pharmacogenetics-based versus conventional dosing of warfarin with promising results^{189,190}. In addition to *VKORC1*, we also identified ASM in *F2* and *F10*, which both are targets for NOACs, and in 11 other hemostatic genes that are current targets of anti-thrombotic, anti-fibrinolytic, or other related drugs. In line with our results, recent studies have reported that DNA methylation may contribute to the variability in response to warfarin¹⁸⁸. Similar findings have been made for aspirin and clopidogrel¹⁹¹. Thus, methylation influenced by genetic variation in *cis* may not only provide a mechanistic link between the non-coding genetic variants and the phenotypic variation observed in circulating hemostatic proteins and prothrombotic diseases, but also in the phenotypic variation in response to drugs.

Overlap between allele-specific expression and DNA methylation (Papers II-III)

Variants regulating mRNA expression in *cis* tend to be found in promoter regions, whereas variants that regulate DNA methylation tend to occur in other regulatory regions such as enhancers and insulators¹⁹², and previous studies have found only a small overlap between variants tagging *cis*-acting expression and DNA methylation.^{20,182,193-196} We thus compared all ASE-SNPs and ASM-SNPs identified in Papers II and III, and found that only 10 SNPs (<1%) were associated with both ASE and ASM. Thus, this outcome is in line with the results of previous studies, and highlights the complimentary nature of conducting both types of mapping.

Correlation of DNA methylation levels between liver and blood (Paper IV)

Given the difficulty in obtaining liver tissue samples, it is of great importance to know whether blood can be used as a proxy for liver for studies of DNA methylation. In Paper IV, we thus used paired liver and blood samples and found that the correlation in methylation between liver and blood was in general very low. When applying a Bonferroni adjusted *P*-value cut-off of less than 0.05, we found significant within-subject correlations in methylation between liver and blood for 354 CpG sites, i.e. 3.4% of all investigated methylation sites. The majority of the correlated CpG sites showed a strong positive correlation ($r > 0.74$ throughout) between methylation levels in blood and liver. Significantly correlated CpGs were over-represented in introns, whereas they were depleted in CGI shores, promoters, exons, exon-intron boundaries, and in the first intron.

For comparison, we also examined the level of CpG methylation correlation between liver and blood across samples. Here, we calculated mean methylation values for each CpG across each tissue, and then calculated the overall correlation of all CpGs between liver and blood. This across-subject correlation was substantially higher (mean $r = 0.68$) compared to the mean within-subject correlation described above.

Two main correlation patterns were identified among the significant CpGs: 1) “continuous” correlation (~24% of the CpGs), where the methylation levels spanned a wide range in a linear fashion, and 2) a trimodal correlation pattern (~47% of CpGs) with methylation levels around 0%, 50% or 100%. The latter

correlation pattern is most likely due to genotype-dependent DNA methylation such as mQTL or ASM. Thus, from a biomarker perspective, the CpGs displaying the trimodal pattern could potentially be assayed using regular DNA sequencing or genotyping, which require less extensive laboratory work and data management in comparison to methylation sequencing.

CpGs displaying variable DNA methylation in liver are more likely to be of biological relevance for the investigated liver-expressed hemostatic genes. Of the 354 CpGs showing a significant correlation between liver and blood, around 90% displayed variable methylation in liver, defined as a methylation range differing at least 5% after exclusion of methylation values above the 90th percentile and below the 10th percentile for each CpG, in liver. Both the trimodal and the continuous group were represented among these. However, we believe that the subset of CpGs that both display variable methylation in liver and a continuous correlation pattern for methylation in liver and blood will be the most useful in the sense of being methylation blood biomarkers for liver. Of all analyzed CpGs, 23% exhibited both these properties. The highest number of CpGs with these properties was found in *F7* (25 CpGs), and all were located in the *F7* promoter and/or in a CGI partly overlapping the promoter. Plasma concentrations of FVII have previously been shown to associate with incident MI and ischemic stroke.¹⁹⁷⁻²⁰⁰ Additionally, methylation of the *F7* promoter in blood (peripheral blood mononuclear cells) has been linked both to plasma concentrations of FVIIa and to CAD.²⁰¹ In light of this, our findings do not only demonstrate that *F7* promoter methylation levels in blood can be used as a surrogate measure for methylation in liver tissue, but also suggest that *F7* promoter methylation in the liver may contribute to the regulation of plasma concentrations of FVII.

CONCLUSIONS TO GIVEN AIMS

Histone acetylation affects the regulation of t-PA mRNA expression in human astrocytes and neurons. Further, expression levels of t-PA are inversely correlated with the degree of DNA methylation in the t-PA promoter. Taken together, our findings suggest involvement of epigenetic mechanisms in t-PA gene regulation in the human brain.

Our study provides the first detailed map of DNA methylation patterns in hemostatic genes in liver. The results show that the promoters of these genes are generally hypomethylated in the liver, as expected for actively transcribed genes. The methylation levels in gene bodies were in general higher, although much more variable.

The investigated hemostatic genes display a high degree of allele-specific expression and DNA methylation in liver. Our findings further demonstrate that analyses of allelic imbalance in expression and methylation are powerful methods to identify putative *cis*-acting variants involved in gene regulation.

In general, DNA methylation levels in the hemostatic genes correlate poorly between liver tissue and peripheral blood within individuals. However, our findings suggest that blood methylation patterns may potentially be used as proxies for liver methylation for a limited subset of CpGs.

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

During the past decade, genome-wide association studies (GWAS) have successfully identified many genetic variants associated with complex diseases and traits. However, these variants typically explain only a small fraction of the estimated heritability. Other mechanisms, such as epigenetics, have been proposed to account for this “missing heritability”. As complex diseases are affected by many different variants with small effects, combining the effects of multiple SNPs with epigenetic information may explain more of the heritability and improve predictive utility of genetic information.

The results presented in this thesis suggest involvement of genetic and epigenetic interactions in the regulation of a set of hemostatic genes predominantly expressed in the liver, and provide evidence for regulatory involvement for a number of novel genetic variants. It may be hypothesized that the combined effects of several SNPs associated with a prothrombotic state might confer risk of thrombotic disorders such as ischemic stroke or venous thromboembolism, whereas SNPs associated with lower plasma concentrations of coagulation factors levels might confer risk for bleeding and hemorrhagic stroke. It would be of great interest to test this hypothesis by constructing a genetic risk score (GRS) based on SNPs identified in this thesis, supplemented with additional SNPs that have been identified through GWAS to associate with circulating levels of prothrombotic factors and/or other relevant phenotypes. The GRS could then be tested for association with related traits, such as ischemic stroke, ischemic stroke subtypes, stroke outcomes, or to VTE, in other available studies. Such analyses could potentially provide further insights into the relative impact of the hemostatic pathway for ischemic stroke and VTE.

Knowledge of tissue-specific DNA methylation provides a foundation for studies of methylation-mediated regulation and tissue diversity in expression. Methylation of CpG dinucleotides are undoubtedly the most abundant and well-studied form of DNA methylation and has been convincingly implicated in chromatin organization and gene regulation. However, DNA methylation may also be found at cytosines in non-CpG contexts, such as CHG and CHH (where “H” corresponds to an A, T or C nucleotide). Non-CpG methylation has been associated with tissue-specific cell functions²⁰² and with transcriptional repression in brain cells²⁰³, and promoter non-CpG methylation has been correlated with gene repression²⁰⁴. However, this phenomenon is not well studied, and the mechanisms and effects of non-CpG methylation are still poorly understood. The research performed within the scope of the present

thesis has focused on CpG methylation. However, the sequence data generated here may in the future also be used for high-resolution investigations of non-CpG methylation in liver and blood in the future.

Genetic and epigenetic mechanisms can affect proteins, phenotypic variation, and expression in various ways. Alternative splicing (AS) is a regulatory mechanism acting on the transcribed RNA which mainly involves “exon skipping”. Individual exons may be included in the final mRNAs under some conditions or in specific tissues, but omitted in others. This process gives rise to multiple mRNA isoforms generated from the same gene, and increases protein diversity significantly. AS is regulated in a cell type-specific manner and is influenced by genetic variation.²⁰⁵ More recently, DNA methylation has also been implicated in AS.²⁰⁶ Most genes are subject to some form of AS, and hemostatic factors are no exception.^{207,208} AS has also been implicated in various diseases²⁰⁹, including stroke²¹⁰ and cardiometabolic disease²¹¹. High-throughput RNA sequencing methods have greatly facilitated analysis and identification of AS, and the deep RNA sequencing data used for the purpose of this theses are well suited for such investigations. In our dataset, information on genotypes and DNA methylation is available from the same individuals, and the effects of these features on hemostatic AS could be comprehensively investigated. Such analyses have the potential to increase our knowledge and understanding of the regulation of hemostatic genes and proteins further.

Three of the four papers included in this thesis are based on targeted gDNA, RNA, and methylation sequencing data. The main advantages of the custom targeted design are that a much higher coverage and sequencing depth for the specific regions of interest can be achieved. For RNA sequencing in particular, whole transcriptome sequencing methods tend to generate much lower coverage for less abundant transcripts. Due to the wide range in expression levels between different mRNA molecules, more abundant transcripts will be sequenced to a higher extent and more reads will be generated for these. Thus, both transcript detection and quantification precision will increase with a deeper sequence level. The extensive sequence depth resulting from targeted RNA sequencing is also useful for identification of novel exons and transcript isoforms that may not be readily detectible in whole RNA-seq data. Furthermore, both expression and DNA methylation vary between different tissues within one individual, as well as between different individuals. Therefore, these types of high-throughput sequencing data are warranted for additional samples, gene sets, and tissue types to fully be able to understand the dynamic nature of biological diversity and to unravel the genetic as well as non-genetic mechanisms involved in gene regulation

ACKNOWLEDGEMENTS - TACK

Jag vill rikta ett varmt tack till alla er som bidragit till denna avhandling och stöttat mig på vägen.

Först och främst vill jag tacka min huvudhandledare *Christina Jern*. Tack för din generositet, uppmuntran och tålamod, och för att du alltid bidrar med nya inblickar och synsätt. Det har varit ovärderligt att få ta del av din stora kunskap din genuina entusiasm kring vår forskning.

Min bihandledare *Tara Stanne*, för ditt stora engagemang i vårt arbete och för ett fantastiskt mentorskap och sällskap på labbet, på konferenser, och i största allmänhet. Det har varit ett privilegium att få arbeta med dig!

Jag vill även tacka mina övriga bihandledare *Marcela Davila Lopez*, thank you for being an amazing teacher and for your positive energy and enthusiasm. Stort tack *Staffan Nilsson* för alla statistikråd såväl som livsråd. Och tack *Lena Hansson* för ovärderlig vägledning genom datalabyrinterna.

Mina nuvarande och tidigare medarbetare i Strokegruppen *Annie Pedersen, Cecilia Lagging, Sofia Klasson, Ingrid Eriksson, Annelie Angerfors, Björn Andersson, Maja Olsson, Erik Lotentzen, Karin Hultman, Ellen Hansson, Anna Tjärnlund-Wolf* samt övriga medarbetare på plan 3, för att ni alltid hjälper till när det behövs och skapar en positiv och kreativ arbetsmiljö. Ett extra tack till *Sofia* för bidraget till omslagsbilden och för din hjälp med allt annat de sista veckorna, och till *Ros-Marie Sjöberg* som håller reda på labbet. *Karin*, tack även för ett fantastiskt handledarskap under min studenttid som lade grunden för ett fortsatt intresse för forskning. Jag vill också rikta ett tack till alla mina före detta medarbetare på CBR, i synnerhet till *Charlotta Blom* för en otroligt kul och lärorik tid.

Ett stort tack till *Christian Blomstrand, Katarina Jood* och alla övriga inblandade i SAHLSIS för trevliga diskussioner, i särskilt under fina vårdagar i Amundövik.

My collaborators at the University Hospital of Geneva, *Egbert Kruithof* and *Sylvie Dunoyer-Geindre*, thank you for letting me visit your lab and for sharing your knowledge, and to *Maurice Curtis* and *Richard Faull* at the University of Auckland, for fruitful collaborations.

Nuvarande och tidigare personal på Genomics och Bioinformatics Core Facility för ovärderlig hjälp i labbet och med analyser. All personal i Leverteamet på Transplantationscentrum vid Sahlgrenska universitetssjukhuset. Jag vill också rikta ett speciellt tack till samtliga patienter som deltagit i våra studier.

Tack till prefekt *Sven Enerbäck* med personal vid Institutionen för Biomedicin, avdelningschef *Göran Larson* vid Avdelningen för Laboratoriemedicin, verksamhetschef *Lars Palmqvist*, sektionschef *Lovisa Lovmar*, och enhetschef *Mirja Henricsson Marcher* vid Klinisk Genetik och Genomik, för att ni skapar och främjar en god forskningsmiljö.

Slutligen vill jag rikta ett stort varmt tack till familj och vänner.

Stefan, min klippa, tack för all kärlek och för det tålamod och stöd du visat, speciellt de senaste veckorna! Och så klart *Alaska*, *Zack* och *Inez* som alltid fyller vardagen med äventyr.

Till min mamma *Elisabeth* och min saknade pappa *Göran*, till mina syskon *Sandra* och *Christian* med familjer, och till *René* och *Antonio*. Tack för uppmuntring och stöd genom livet.

Till mina svärföräldrar *Annelie* och *Tommy*. Det har varit ovärderligt att få andas ut på Ålekvarn emellanåt under arbetet med denna avhandling.

Stort tack till mina fantastiska vänner, i synnerhet *Stina*, *Rebecca* och *Daniel*, för att ni alltid har tid att lyssna, stötta, och dela med er av stora som små saker.

The work described in this thesis was supported by the Swedish Research Council, the Swedish Heart and Lung Foundation, the Swedish State under the agreement between the Swedish Government and the County Councils (the ALF-agreement), the Swedish Stroke Association, the Gothenburg Foundation for Neurological Research, the Swedish Foundation for Strategic Research, the Wallenberg Advanced Bioinformatics Infrastructure (WABI), the Rune and Ulla Amlöv Foundation, the John and Brit Wennerström Foundation, the Marcus Borgström Foundation, and the Nilsson-Ehle Endowments.

REFERENCES

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. 4 ed. *New York: Garland Science*; 2002.
2. Watson JD, Crick FH. Molecular structure of nucleic acids; A structure for deoxyribose nucleic acid. *Nature*. 1953;171:737-738.
3. Fischle W, Wang Y, Allis CD. Histone and chromatin cross-talk. *Curr Opin Cell Biol*. 2003;15:172-183.
4. Smith CL, Peterson CL. ATP-dependent chromatin remodeling. *Curr Top Dev Biol*. 2005;65:115-148.
5. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431:931-945.
6. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291:1304-1351.
7. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56-65.
8. Robert F, Pelletier J. Exploring the impact of single-nucleotide polymorphisms on translation. *Front Genet*. 2018;9:507.
9. Goldstein DB, Weale ME. Population genomics: Linkage disequilibrium holds the key. *Curr Biol*. 2001;11:R576-579.
10. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*. 2007;447:425-432.
11. Weber M, Schubeler D. Genomic patterns of DNA methylation: Targets and function of an epigenetic mark. *Curr Opin Cell Biol*. 2007;19:273-280.
12. Foley DL, Craig JM, Morley R, et al. Prospects for epigenetic epidemiology. *Am J Epidemiol*. 2009;169:389-400.
13. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921.
14. Strichman-Almashanu LZ, Lee RS, Onyango PO, et al. A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome Res*. 2002;12:543-554.
15. Robertson KD. DNA methylation and human disease. *Nat Rev Genet*. 2005;6:597-610.
16. Brenet F, Moh M, Funk P, et al. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One*. 2011;6:e14524.
17. Jjinglo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. On the presence and role of human gene-body DNA methylation. *Oncotarget*. 2012;3:462-474.
18. Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462:315-322.
19. Goto T, Monk M. Regulation of X-chromosome inactivation in development in mice and humans. *Microbiol Mol Biol Rev*. 1998;62:362-378.
20. Do C, Lang CF, Lin J, et al. Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. *Am J Hum Genet*. 2016;98:934-955.
21. Marzi SJ, Meaburn EL, Dempster EL, et al. Tissue-specific patterns of allelically-skewed DNA methylation. *Epigenetics*. 2016;11:24-35.
22. Cedar H, Bergman Y. Linking DNA methylation and histone modification: Patterns and paradigms. *Nat Rev Genet*. 2009;10:295-304.
23. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001;293:1074-1080.
24. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000;403:41-45.

25. Fuks F. DNA methylation and histone modifications: Teaming up to silence genes. *Curr Opin Genet Dev.* 2005;15:490-495.
26. Franklin RE, Gosling RG. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature.* 1953;172:156-157.
27. Wilkins MH, Randall JT. Crystallinity in sperm heads: Molecular structure of nucleoprotein in vivo. *Biochim Biophys Acta.* 1953;10:192-193.
28. Holley RW, Apgar J, Everett GA, et al. Structure of a ribonucleic acid. *Science.* 1965;147:1462-1465.
29. Holley RWM, James T, Zamir A. A new method for sequence determination of large oligonucleotides. *Biochem Biophys Res Commun.* 1964;17:389-394.
30. Madison JT, Holley RW. The presence of 5,6-dihydrouridylic acid in yeast "soluble" ribonucleic acid. *Biochem Biophys Res Commun.* 1965;18:153-157.
31. Sanger F, Brownlee GG, Barrell BG. A two-dimensional fractionation procedure for radioactive nucleotides. *J Mol Biol.* 1965;13:373-398.
32. Padmanabhan R, Jay E, Wu R. Chemical synthesis of a primer and its use in the sequence analysis of the lysozyme gene of bacteriophage T4. *Proc Natl Acad Sci USA.* 1974;71:2510-2514.
33. Padmanabhan R, Wu R. Nucleotide sequence analysis of DNA. IX. Use of oligonucleotides of defined sequence as primers in DNA sequence analysis. *Biochem Biophys Res Commun.* 1972;48:1295-1302.
34. Sanger F, Donelson JE, Coulson AR, Kossel H, Fischer D. Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage ϕ 1 DNA. *Proc Natl Acad Sci USA.* 1973;70:1209-1213.
35. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA.* 1977;74:560-564.
36. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 1975;94:441-448.
37. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA.* 1977;74:5463-5467.
38. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harb Symp Quant Biol.* 1986;51:263-273.
39. Nyren P, Pettersson B, Uhlen M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal Biochem.* 1993;208:171-175.
40. Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science.* 1998;281:363-365.
41. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol.* 2008;26:1135-1145.
42. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17:333-351.
43. Head SR, Komori HK, LaMere SA, et al. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques.* 2014;56:61-64.
44. Quail MA, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 2012;13:341.
45. Dewey FE, Grove ME, Pan C, et al. Clinical interpretation and implications of whole-genome sequencing. *JAMA.* 2014;311:1035-1045.
46. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: The next big thing in genetic medicine. *American J Hum Genet.* 2009;85:142-154.
47. Waterman M, Uberbacher E, Spengler S, et al. Genome informatics I: Community databases. *J Comput Biol.* 1994;1:173-190.

48. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Updated 2019. Accessed 16 December, 2019.
49. Venter JC, Smith HO, Adams MD. The Sequence of the human genome. *Clin Chem.* 2015;61:1207-1208.
50. International HapMap Consortium. The International HapMap Project. *Nature.* 2003;426:789-796.
51. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010;467:52-58.
52. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science.* 2004;306:636-640.
53. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467:1061-1073.
54. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature.* 2015;526:68-74.
55. Peplow M. The 100,000 Genomes Project. *BMJ.* 2016;353:i1757.
56. Turnbull C, Scott RH, Thomas E, et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ.* 2018;361:k1687.
57. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *Hum Mol Genet.* 2018;27:3641-3649.
58. Evangelou E, Warren HR, Mosen-Ansorena D, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature genetics.* 2018;50:1412-1425.
59. 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:e0144997.
60. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet.* 2018;50:524-537.
61. Scott RA, Scott LJ, Magi R, et al. An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes.* 2017;66:2888-2902.
62. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet.* 2013;45:1452-1458.
63. Söderholm M, Pedersen A, Lorentzen E, et al. Genome-wide association meta-analysis of functional outcome after ischemic stroke. *Neurology.* 2019;92:e1271-e1283.
64. Giacomini KM, Yee SW, Mushiroda T, Weinshilboum RM, Ratain MJ, Kubo M. Genome-wide association studies of drug response and toxicity: An opportunity for genome medicine. *Nat Rev Drug Discov.* 2017;16:1.
65. Rockman MV, Kruglyak L. Genetics of global gene expression. *Nat Rev Genet.* 2006;7:862-872.
66. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45:580-585.
67. Hannon E, Gorrie-Stone TJ, Smart MC, et al. Leveraging DNA-methylation quantitative-trait loci to characterize the relationship between methylomic variation, gene expression, and complex traits. *Am J Hum Genet.* 2018;103:654-665.
68. Knight JC. Allele-specific gene expression uncovered. *Trends Genet.* 2004;20:113-116.
69. Hultman K, Tjarnlund-Wolf A, Odeberg J, Eriksson P, Jern C. Allele-specific transcription of the PAI-1 gene in human astrocytes. *Thromb Haemost.* 2010;104:998-1008.

70. Tjarnlund-Wolf A, Hultman K, Curtis MA, Faull RL, Medcalf RL, Jern C. Allelic imbalance of tissue-type plasminogen activator (t-PA) gene expression in human brain tissue. *Thromb Haemost.* 2011;105:945-953.
71. Eriksson P, Kallin B, van 't Hooft FM, Bavenholm P, Hamsten A. Allele-specific increase in basal transcription of the plasminogen-activator inhibitor 1 gene is associated with myocardial infarction. *Proc Natl Acad Sci U S A.* 1995;92:1851-1855.
72. Pastinen T. Genome-wide allele-specific analysis: Insights into regulatory variation. *Nat Rev Genet.* 2010;11:533-538.
73. Tycko B. Allele-specific DNA methylation: Beyond imprinting. *Hum Mol Genet.* 2010;19:R210-220.
74. Jackson SP. The growing complexity of platelet aggregation. *Blood.* 2007;109:5087-5095.
75. Varga-Szabo D, Pleines I, Nieswandt B. Cell adhesion mechanisms in platelets. *Arterioscler Thromb Vasc Biol.* 2008;28:403-412.
76. Dahlbäck B. Blood coagulation. *Lancet.* 2000;355:1627-1632.
77. Furie B. Pathogenesis of thrombosis. *Hematology Am Soc Hematol Educ Program.* 2009:255-258.
78. Kahn ML, Zheng YW, Huang W, et al. A dual thrombin receptor system for platelet activation. *Nature.* 1998;394:690-694.
79. Cramer TJ, Griffin JH, Gale AJ. Factor V is an anticoagulant cofactor for activated protein C during inactivation of factor Va. *Pathophysiol Haemost Thromb.* 2010;37:17-23.
80. Esmon CT, Owen WG. Identification of an endothelial cell cofactor for thrombin-catalyzed activation of protein C. *Proc Natl Acad Sci U S A.* 1981;78:2249-2252.
81. Rau JC, Beaulieu LM, Huntington JA, Church FC. Serpins in thrombosis, hemostasis and fibrinolysis. *J Thromb Haemost.* 2007;5 Suppl 1:102-115.
82. Rijken DC, Lijnen HR. New insights into the molecular mechanisms of the fibrinolytic system. *J Thromb Haemost.* 2009;7:4-13.
83. Longstaff C, Thelwell C, Williams SC, Silva MM, Szabo L, Kolev K. The interplay between tissue plasminogen activator domains and fibrin structures in the regulation of fibrinolysis: kinetic and microscopic studies. *Blood.* 2011;117:661-668.
84. Marx PF, Dawson PE, Bouma BN, Meijers JC. Plasmin-mediated activation and inactivation of thrombin-activatable fibrinolysis inhibitor. *Biochemistry.* 2002;41:6688-6696.
85. Dashty M, Akbarkhanzadeh V, Zeebregts CJ, et al. Characterization of coagulation factor synthesis in nine human primary cell types. *Sci Rep.* 2012;2:787.
86. Lisman T, Caldwell SH, Burroughs AK, et al. Hemostasis and thrombosis in patients with liver disease: The ups and downs. *J Hepatol.* 2010;53:362-371.
87. Bolton-Maggs PH, Pasi KJ. Haemophilias A and B. *Lancet.* 2003;361:1801-1809.
88. Franchini M, Mannucci PM. Inhibitors of propagation of coagulation (factors VIII, IX and XI): A review of current therapeutic practice. *Br J Clin Pharmacol.* 2011;72:553-562.
89. Kunicki TJ, Federici AB, Salomon DR, et al. An association of candidate gene haplotypes and bleeding severity in von Willebrand disease (VWD) type 1 pedigrees. *Blood.* 2004;104:2359-2367.
90. Mulder R, Croles FN, Mulder AB, Huntington JA, Meijer K, Lukens MV. SERPINC1 gene mutations in antithrombin deficiency. *Br J Haematol.* 2017;178:279-285.
91. Gomez K, Laffan MA. Hunting for the mutation in inherited thrombophilia. *Blood Coagul Fibrinolysis.* 2004;15:125-127.

92. Garcia de Frutos P, Fuentes-Prior P, Hurtado B, Sala N. Molecular basis of protein S deficiency. *Thromb Haemost.* 2007;98:543-556.
93. Tiscia GL, Margaglione M. Human Fibrinogen: Molecular and genetic aspects of congenital disorders. *Int J Mol Sci.* 2018;19:E1597
94. Akhavan S, Mannucci PM, Lak M, et al. Identification and three-dimensional structural analysis of nine novel mutations in patients with prothrombin deficiency. *Thromb Haemost.* 2000;84:989-997.
95. Rosendaal FR, Reitsma PH. Genetics of venous thrombosis. *J Thromb Haemost.* 2009;7 Suppl 1:301-304.
96. Bertina RM, Koeleman BP, Koster T, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature.* 1994;369:64-67.
97. Cutler JA, Patel R, Rangarajan S, Tait RC, Mitchell MJ. Molecular characterization of 11 novel mutations in patients with heterozygous and homozygous FV deficiency. *Haemophilia.* 2010;16:937-942.
98. Dawson SJ, Wiman B, Hamsten A, Green F, Humphries S, Henney AM. The two allele sequences of a common polymorphism in the promoter of the plasminogen activator inhibitor-1 (PAI-1) gene respond differently to interleukin-1 in HepG2 cells. *J Biol Chem.* 1993;268:10739-10745.
99. Green F, Hamsten A, Blombäck M, Humphries S. The role of beta-fibrinogen genotype in determining plasma fibrinogen levels in young survivors of myocardial infarction and healthy controls from Sweden. *Thromb Haemost.* 1993;70:915-920.
100. Scarabin PY, Bara L, Ricard S, et al. Genetic variation at the beta-fibrinogen locus in relation to plasma fibrinogen concentrations and risk of myocardial infarction. The ECTIM Study. *Arterioscler Thromb.* 1993;13:886-891.
101. Thomas AE, Green FR, Kelleher CH, et al. Variation in the promoter region of the beta fibrinogen gene is associated with plasma fibrinogen levels in smokers and non-smokers. *Thromb Haemost.* 1991;65:487-490.
102. van 't Hooft FM, von Bahr SJ, Silveira A, Iliadou A, Eriksson P, Hamsten A. Two common, functional polymorphisms in the promoter region of the beta-fibrinogen gene contribute to regulation of plasma fibrinogen concentration. *Arterioscler Thromb Vasc Biol.* 1999;19:3063-3070.
103. van 't Hooft FM, Silveira A, Tomvall P, et al. Two common functional polymorphisms in the promoter region of the coagulation factor VII gene determining plasma factor VII activity and mass concentration. *Blood.* 1999;93:3432-3441.
104. 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;25:358-370.
105. Sabater-Lleal M, Huang J, Chasman D, et al. Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation.* 2013;128:1310-1324.
106. Olson NC, Raffield LM, Lange LA, et al. Associations of activated coagulation factor VII and factor VIIa-antithrombin levels with genome-wide polymorphisms and cardiovascular disease risk. *J Thromb Haemost.* 2018;16:19-30.
107. Smith NL, Chen MH, Dehghan A, et al. 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;121:1382-1392.
108. Sennblad B, Basu S, Mazur J, et al. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor XI levels. *Hum Mol Genet.* 2017;26:637-649.

109. 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;118:298-308.
110. 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;16:2024-2034.
111. Freson K, Izzi B, Van Geet C. From genetics to epigenetics in platelet research. *Thromb Res.* 2012;129:325-329.
112. Rakyán VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet.* 2011;12:529-541.
113. Davis Armstrong NM, Chen WM, Brewer MS, et al. Epigenome-wide analyses identify two novel associations with recurrent stroke in the Vitamin Intervention for Stroke Prevention Clinical trial. *Front Genet.* 2018;9:358.
114. Krupinski J, Carrera A, Muino E, et al. DNA Methylation in stroke. Update of latest advances. *Comput Struct Biotechnol J.* 2018;16:1-5.
115. Nakatochi M, Ichihara S, Yamamoto K, et al. Epigenome-wide association of myocardial infarction with DNA methylation sites at loci related to cardiovascular disease. *Clin Epigenetics.* 2017;9:54.
116. Rask-Andersen M, Martinsson D, Ahsan M, et al. Epigenome-wide association study reveals differential DNA methylation in individuals with a history of myocardial infarction. *Hum Mol Genet.* 2016;25:4739-4748.
117. Fernandez-Sanles A, Sayols-Baixeras S, Curcio S, Subirana I, Marrugat J, Elosua R. DNA methylation and age-independent cardiovascular risk, an epigenome-wide approach: The REGICOR Study (REgistre GIroni del COR). *Arterioscler Thromb Vasc Biol.* 2018;38:645-652.
118. Agha G, Mendelson MM, Ward-Caviness CK, et al. Blood leukocyte DNA methylation predicts risk of future myocardial infarction and coronary heart disease. *Circulation.* 2019;140:645-657.
119. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47:1121-1130.
120. Previtali E, Bucciarelli P, Passamonti SM, Martinelli I. Risk factors for venous and arterial thrombosis. *Blood Transfus.* 2011;9:120-138.
121. Kannel WB, Wolf PA, Castelli WP, D'Agostino RB. Fibrinogen and risk of cardiovascular disease. The Framingham Study. *JAMA.* 1987;258:1183-1186.
122. Rothwell PM, Howard SC, Power DA, et al. Fibrinogen concentration and risk of ischemic stroke and acute coronary events in 5113 patients with transient ischemic attack and minor ischemic stroke. *Stroke.* 2004;35:2300-2305.
123. Iacoviello L, Di Castelnuovo A, De Knijff P, et al. Polymorphisms in the coagulation factor VII gene and the risk of myocardial infarction. *N Engl J Med.* 1998;338:79-85.
124. Lopaciuk S, Windyga J, Watala CW, et al. Polymorphisms in the factor VII gene and ischemic stroke in young adults. *Blood Coagul Fibrinolysis.* 2010;21:442-447.
125. Rudnicka AR, Mt-Isa S, Meade TW. Associations of plasma fibrinogen and factor VII clotting activity with coronary heart disease and stroke: prospective cohort study from the screening phase of the Thrombosis Prevention Trial. *J Thromb Haemost.* 2006;4:2405-2410.
126. Catto AJ, Carter AM, Stickland M, Bamford JM, Davies JA, Grant PJ. Plasminogen activator inhibitor-1 (PAI-1) 4G/5G promoter polymorphism and levels in subjects with cerebrovascular disease. *Thromb Haemost.* 1997;77:730-734.
127. Hamsten A, de Faire U, Walldius G, et al. Plasminogen activator inhibitor in plasma: Risk factor for recurrent myocardial infarction. *Lancet.* 1987;2:3-9.

128. Hamsten A, Wiman B, de Faire U, Blomback M. Increased plasma levels of a rapid inhibitor of tissue plasminogen activator in young survivors of myocardial infarction. *N Engl J Med.* 1985;313:1557-1563.
129. Lindgren A, Lindoff C, Norrving B, Astedt B, Johansson BB. Tissue plasminogen activator and plasminogen activator inhibitor-1 in stroke patients. *Stroke.* 1996;27:1066-1071.
130. Liu Y, Cheng J, Guo X, et al. The roles of PAI-1 gene polymorphisms in atherosclerotic diseases: A systematic review and meta-analysis involving 149,908 subjects. *Gene.* 2018;673:167-173.
131. Andersson HM, Siegerink B, Luken BM, et al. High VWF, low ADAMTS13, and oral contraceptives increase the risk of ischemic stroke and myocardial infarction in young women. *Blood.* 2012;119:1555-1560.
132. Andersson J, Johansson L, Ladenvall P, et al. C-reactive protein is a determinant of first-ever stroke: prospective nested case-referent study. *Cerebrovasc Dis.* 2009;27:544-551.
133. Chang TR, Albright KC, Boehme AK, et al. Factor VIII in the setting of acute ischemic stroke among patients with suspected hypercoagulable state. *Clin Appl Thromb Hemost.* 2014;20:124-128.
134. Folsom AR, Rosamond WD, Shahar E, et al. Prospective study of markers of hemostatic function with risk of ischemic stroke. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. *Circulation.* 1999;100:736-742.
135. Hanson E, Jood K, Karlsson S, Nilsson S, Blomstrand C, Jern C. Plasma levels of von Willebrand factor in the etiologic subtypes of ischemic stroke. *J Thromb Haemost.* 2011;9:275-281.
136. Hanson E, Kanse SM, Joshi A, et al. Plasma factor VII-activating protease antigen levels and activity are increased in ischemic stroke. *J Thromb Haemost.* 2012;10:848-856.
137. Stanne TM, Hanson E, Olsson S, et al. Factor VII antigen levels are differentially associated to etiologic subtypes of ischaemic stroke. *Thromb Haemost.* 2013;110:1305-1306.
138. Ekstrom M, Silveira A, Bennermo M, Eriksson P, Tornvall P. Coagulation factor VII and inflammatory markers in patients with coronary heart disease. *Blood Coagul Fibrinolysis.* 2007;18:473-477.
139. Folkersen L, Fauman E, Sabater-Lleal M, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS genetics.* 2017;13:e1006706.
140. Fisher MJ. Brain regulation of thrombosis and hemostasis: from theory to practice. *Stroke.* 2013;44:3275-3285.
141. Bajaj MS, Kuppuswamy MN, Manepalli AN, Bajaj SP. Transcriptional expression of tissue factor pathway inhibitor, thrombomodulin and von Willebrand factor in normal human tissues. *Thromb Haemost.* 1999;82:1047-1052.
142. Wong VL, Hofman FM, Ishii H, Fisher M. Regional distribution of thrombomodulin in human brain. *Brain Res.* 1991;556:1-5.
143. Drake TA, Morrissey JH, Edgington TS. Selective cellular expression of tissue factor in human tissues. Implications for disorders of hemostasis and thrombosis. *Am J Pathol.* 1989;134:1087-1097.
144. Eddleston M, de la Torre JC, Oldstone MB, Loskutoff DJ, Edgington TS, Mackman N. Astrocytes are the primary source of tissue factor in the murine central nervous system. A role for astrocytes in cerebral hemostasis. *J Clin Invest.* 1993;92:349-358.
145. Melchor JP, Strickland S. Tissue plasminogen activator in central nervous system physiology and pathology. *Thromb Haemost.* 2005;93:655-660.

146. Teesalu T, Kulla A, Simisker A, et al. Tissue plasminogen activator and neuroserpin are widely expressed in the human central nervous system. *Thromb Haemost.* 2004;92:358-368.
147. Qian Z, Gilbert ME, Colicos MA, Kandel ER, Kuhl D. Tissue-plasminogen activator is induced as an immediate-early gene during seizure, kindling and long-term potentiation. *Nature.* 1993;361:453-457.
148. Seeds NW, Basham ME, Ferguson JE. Absence of tissue plasminogen activator gene or activity impairs mouse cerebellar motor learning. *J Neurosci.* 2003;23:7368-7375.
149. Fernandez-Monreal M, Lopez-Atalaya JP, Benchenane K, et al. Is tissue-type plasminogen activator a neuromodulator? *Mol Cell Neurosci.* 2004;25:594-601.
150. Alonso M, Medina JH, Pozzo-Miller L. ERK1/2 activation is necessary for BDNF to increase dendritic spine density in hippocampal CA1 pyramidal neurons. *Learn Mem.* 2004;11:172-178.
151. Qi M, Zhuo M, Skalhegg BS, et al. Impaired hippocampal plasticity in mice lacking the Cbeta1 catalytic subunit of cAMP-dependent protein kinase. *Proc Natl Acad Sci U S A.* 1996;93:1571-1576.
152. Zhuo M, Holtzman DM, Li Y, et al. Role of tissue plasminogen activator receptor LRP in hippocampal long-term potentiation. *J Neurosci.* 2000;20:542-549.
153. Pang PT, Teng HK, Zaitsev E, et al. Cleavage of proBDNF by tPA/plasmin is essential for long-term hippocampal plasticity. *Science.* 2004;306:487-491.
154. Nicole O, Docagne F, Ali C, et al. The proteolytic activity of tissue-plasminogen activator enhances NMDA receptor-mediated signaling. *Nat Med.* 2001;7:59-64.
155. Yepes M, Sandkvist M, Moore EG, Bugge TH, Strickland DK, Lawrence DA. Tissue-type plasminogen activator induces opening of the blood-brain barrier via the LDL receptor-related protein. *J Clin Invest.* 2003;112:1533-1540.
156. Medcalf RL. Fibrinolysis: From blood to the brain. *J Thromb Haemost.* 2017;15:2089-2098.
157. von Bartheld CS, Bahney J, Herculano-Houzel S. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *J Comp Neurol.* 2016;524:3865-3895.
158. Burda JE, Bernstein AM, Sofroniew MV. Astrocyte roles in traumatic brain injury. *Exp Neurol.* 2016;275:305-315.
159. Pan C, Kumar C, Bohl S, Klingmueller U, Mann M. Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol Cell Proteomics.* 2009;8:443-450.
160. Brodal P. The Central Nervous System. 4 ed. *Oxford University Press*; Oxford 2010.
161. Tajadini M, Panjehpour M, Javanmard SH. Comparison of SYBR Green and TaqMan methods in quantitative real-time polymerase chain reaction analysis of four adenosine receptor subtypes. *Adv Biomed Res.* 2014;3:85.
162. Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis. *Genome Biology.* 2016;17:13.
163. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28:27-30.
164. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1:417-425.
165. 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 Genetics.* 2011;7:e1002078.
166. 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;39:226-231.

167. Sun YV. The influences of genetic and environmental factors on methylome-wide association studies for human diseases. *Curr Genet Med Rep.* 2014;2:261-270.
168. Kurdyukov S, Bullock M. DNA methylation analysis: choosing the right method. *Biology (Basel).* 2016;5.
169. Degner JF, Marioni JC, Pai AA, et al. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics (Oxford, England).* 2009;25:3207-3212.
170. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29:308-311.
171. Hubbard T, Barker D, Birney E, et al. The Ensembl genome database project. *Nucleic Acids Res.* 2002;30:38-41.
172. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42:D1001-1006.
173. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: A database of human genotype-phenotype associations. *Bioinformatics (Oxford, England).* 2016;32:3207-3209.
174. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res.* 2002;12:996-1006.
175. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40:D930-934.
176. Dunoyer-Geindre S, Kruihof EK. Epigenetic control of tissue-type plasminogen activator synthesis in human endothelial cells. *Cardiovasc Res.* 2011;90:457-463.
177. Kooistra T, Schrauwen Y, Arts J, Emeis JJ. Regulation of endothelial cell t-PA synthesis and release. *Int J Hematol.* 1994;59:233-255.
178. Jood K, Ladenvall P, Tjarnlund-Wolf A, et al. Fibrinolytic gene polymorphism and ischemic stroke. *Stroke.* 2005;36:2077-2081.
179. Gutierrez-Arcelus M, Ongen H, Lappalainen T, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLoS Genetics.* 2015;11:e1004958.
180. 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.
181. Hannon E, Spiers H, Viana J, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci.* 2016;19:48-54.
182. 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:R37.
183. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet.* 2011;43:339-344.
184. Ikram MA, Seshadri S, Bis JC, et al. Genomewide association studies of stroke. *N Engl J Med.* 2009;360:1718-1728.
185. Teichert M, Eijgelsheim M, Uitterlinden AG, et al. Dependency of phenprocoumon dosage on polymorphisms in the VKORC1, CYP2C9, and CYP4F2 genes. *Pharmacogenet Genom.* 2011;21:26-34.
186. Cullell N, Carrera C, Muino 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;9:29238-29258.
187. Kimmel SE, French B, Kasner SE, et al. A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med.* 2013;369:2283-2293.

188. Luo Z, Liu R, Sun B, et al. Identification of gene modules associated with warfarin dosage by a genome-wide DNA methylation study. *Die Pharmazie*. 2018;73:288-293.
189. Pirmohamed M, Burnside G, Eriksson N, et al. A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med*. 2013;369:2294-2303.
190. Shi C, Yan W, Wang G, Wang F, Li Q, Lin N. Pharmacogenetics-based versus conventional dosing of warfarin: A meta-analysis of randomized controlled trials. *PLoS One*. 2015;10:e0144511.
191. Gallego-Fabrega C, Carrera C, Reny JL, et al. PPM1A methylation is associated with vascular recurrence in aspirin-treated patients. *Stroke*. 2016;47:1926-1929.
192. Do C, Shearer A, Suzuki M, et al. Genetic-epigenetic interactions in *cis*: a major focus in the post-GWAS era. *Genome biology*. 2017;18:120.
193. Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome biology*. 2016;17:61.
194. Grundberg E, Meduri E, Sandling JK, et al. 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;93:876-890.
195. 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:e0157776.
196. 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;18:340-346.
197. Folsom AR, Wu KK, Rosamond WD, Sharrett AR, Chambless LE. Prospective study of hemostatic factors and incidence of coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) Study. *Circulation*. 1997;96:1102-1108.
198. Meade TW, Mellows S, Brozovic M, et al. Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. *Lancet*. 1986;2:533-537.
199. Tracy RP, Arnold AM, Ettinger W, Fried L, Meilahn E, Savage P. The relationship of fibrinogen and factors VII and VIII to incident cardiovascular disease and death in the elderly: results from the cardiovascular health study. *Arterioscler Thromb Vasc Biol*. 1999;19:1776-1783.
200. Zakai NA, Lange L, Longstreth WT, Jr., et al. Association of coagulation-related and inflammation-related genes and factor VIIc levels with stroke: the Cardiovascular Health Study. *J Thromb Haemost*. 2011;9:267-274.
201. Friso S, Lotto V, Choi SW, et al. Promoter methylation in coagulation F7 gene influences plasma FVII concentrations and relates to coronary artery disease. *J Med Genet*. 2012;49:192-199.
202. Schultz MD, He Y, Whitaker JW, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*. 2015;523:212-216.
203. Guo JU, Su Y, Shin JH, et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci*. 2014;17:215-222.
204. Inoue S, Oishi M. Effects of methylation of non-CpG sequence in the promoter region on the expression of human synaptotagmin XI (*syt11*). *Gene*. 2005;348:123-134.
205. Estivill X. Genetic variation and alternative splicing. *Nat Biotechnol*. 2015;33:357-359.
206. Shayeveitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. *RNA*. 2018;24:1351-1362.

207. Bogdanov VY. Blood coagulation and alternative pre-mRNA splicing: An overview. *Curr Mol Med.* 2006;6:859-869.
208. Dahlback B. Novel insights into the regulation of coagulation by factor V isoforms, tissue factor pathway inhibitor α , and protein S. *J Thromb Haemost.* 2017;15:1241-1250.
209. Ward AJ, Cooper TA. The pathobiology of splicing. *J Pathol.* 2010;220:152-163.
210. Dykstra-Aiello C, Jickling GC, Ander BP, et al. Intracerebral hemorrhage and ischemic stroke of different etiologies have distinct alternatively spliced mRNA profiles in the blood: A pilot RNA-seq study. *Transl Stroke Res.* 2015;6:284-289.
211. Franzen O, Ermel R, Cohain A, et al. Cardiometabolic risk loci share downstream *cis*- and *trans*-gene regulation across tissues and diseases. *Science.* 2016;353:827-830.