CHAPTER II

MATERIALS AND METHODS

2.1 PATIENTS

2.1.1 Uppsala study

In Uppsala University Hospital, 201 patients who were treated with warfarin at the anticoagulation clinic were recruited in 2000. For six consecutive visits in the clinics, five weekly warfarin doses and the corresponding prothrombin time international normalised ratio (PT INR) values were registered. The interval of each visit for each patient ranges between 1 week and 1 month, depending on their medication. The 5 weekly doses were averaged for subsequent analyses, and 166 patients were stable (defined as 3 consistent doses) at the time of recruitment.

The individual warfarin dose varied nearly 20 fold, ranging from 4.5 to 77.25 mg per week. Information such as age, gender, bodyweight (BW), other diseases and indication for treatment was retrieved from the patients' medical records although there are seven patients whose body weight details are unknown. Body weights were obtained from medical records but these values may have varied during the course of warfarin treatment as these were not recorded at each visit.

The patients were mostly Caucasians: 194 were of Swedish origin; four were of European descent and three were from the Middle East. These patients were 28-88 years old (median: 69) when their blood samples were collected, and they had been treated with warfarin for at least 2 months, ranging from 2.4 months to 26 years. These patients were grouped by their treatment indications: patients with heart valve prosthesis which aims to target a higher PT

INR value and patients treated for other indications which aims to a normal 2.0 - 3.0 PT INR value.

The concurrent medications were also recorded and checked according to the MICROMEDEX database (http://www.micromedex.com/). The drugs were categorised as interacting if they had been described to have moderate or major interactions with warfarin. In 201 warfarin patients, there are 107 concurrent medications being prescribed which were known to influence the warfarin therapy. For further analysis in concurrent medication, patients are classified into three groups: individuals requiring high warfarin doses due to interaction with other concurrent prescription; individuals requiring lower warfarin doses because the medications enhance the effect of warfarin and patients without any known interactions.

2.1.2 WARG study

A larger Swedish cohort was used for validating any finding in the Uppsala cohort. The patients were recruited through the Warfarin Genetics (WARG) project: a nation-wide, prospective case-control in warfarin treated patients involving 40 Swedish centres, and coordinated by the Karolinska Institute in Sweden. Patients aged less than 18 years old and those previously treated with warfarin were excluded from this study. Apart from these and established contraindications to warfarin treatment (e.g. pregnancy), this study has no other exclusion criteria. A total number of 1523 patients were recruited and 1496 samples were analyzed. These patients were 18-92 years old (median: 66) when they were first treated with warfarin, and the individual warfarin doses varied also nearly 20 fold, ranging from 6.00 to 113.65 mg per week.

In order to not to interfere with the local management of warfarin treatment, warfarin dose and PT INR measurement intervals are chosen at the discretion of the treating physician. The concurrent medication was recorded and classified according to its potential for pharmacologic interactions with warfarin, by use of the classification presented in the classifications in the Swedish Drug Index (Läkemedelsindustriföreningen 2007), the drug information databases Janusinfo (www.janusinfo.se), and Micromedex (www.thomsonhc.com). Dosages, INR-values, concomitant medication and complications are continuously recorded in the Internet-based medical record system.

2.1.3 Patients subject to severe bleeding

Twelve patients out of the 201 in the Uppsala study were subject to one or more severe bleeding episodes and an additional 24 bleeding patients were recruited from other anticoagulation clinics in Sweden by Dr Mia Wadelius. Severe bleeding is defined as bleeding causing hospitalisation for at least one night but excluding the following: bleeding induced by thrombolysis, post-surgical operation, trauma, and malignancy, and bleeding in patients aged >88 years old. Thus, a total of 36 DNA samples from bleeding patients were sent from Uppsala University. In the WARG study, 28 out of 1523 patients had severe bleeding. In total, 64 DNA samples were collected from warfarin bleeders.

2.1.4 DNA preparation

Patient sample DNA was extracted from whole blood in Uppsala University Hospital (for Uppsala study) and the Karolinska Intitute (for WARG study) in Sweden. The methods and kits for DNA extraction were different in the two institutions (details given below). The

purified DNAs were frozen and shipped to the Sanger Institute on dry ice. When the DNA samples arrived, the DNA was defrosted, centrifuged at 4000 rpm for 5 minutes and then stored at -20°C.

2.1.4.1 Uppsala study

The blood sample was collected in ethylenediaminetetraacetic acid (EDTA)-anticoagulated tube and DNA was extracted using QIAamp DNA Blood Mini Kit (Qiagen). In a reaction, 200 µl EDTA-treated blood, 20 µl QIAgen protease and 200 µl Al buffer (lysis buffer) were mixed in a microcentrifuge tube and vortexed for 15 seconds. The tube was then incubated at 56°C for 10 minutes, followed by a brief centrifugation to remove drops from the inside of the lid. 200 µl of 100% ethanol was then added and the tube vortexed for 15 seconds. The mix was then transferred to the QIAamp Spin Column and centrifuged at 6000g for 1 minute. 500 µl AW1 buffer (wash buffer) was added and centrifuged at 6000g for 1 minute. After centrifugation, 500 µl AW2 buffer (wash buffer) was added and centrifuged at 20000g for 3 minutes. The columns were then placed in Eppendorf tubes and 200 µl AE buffer (elution buffer) was added. The samples were incubated for 1-5 minutes at room temperature before being centrifuged at 6000g for 1 minute.

2.1.4.2 WARG study

DNA was prepared from 1 ml EDTA-treated whole blood using the MagnaPure LC method according to the manufacturer's instructions (MagnaPure DNA Isolation Kit-Large Volume; Roche Diagnostics, Mannheim, Germany).

2.2 GENETIC MARKER SELECTION

2.2.1 Single nucleotide polymorphism (SNP)

SNPs in the candidate genes were selected from Ensembl SNP database (http://www.ensembl.org/) with the aim of at least 5 kb spacing between them. Functional variants on candidate cytochrome P450 genes were tested in the Uppsala study according to the annotation in Allele nomenclature for Cytochrome P450 enzymes website (http://www.cypalleles.ki.se/).

Validated SNPs with minor allele frequency (MAF) greater than 5% were analysed to select tag SNPs using Tagger (section 2.6.2). These tag SNPs, together with SNPs found to be nominally significant in Uppsala study, were then tested in WARG study.

2.2.2 Microsatellite repeat marker

The microsatellite marker in intron 6 of the *GGCX* gene has been reported to be associated with warfarin dose among individuals (Shikata et al. 2004) and was selected for genotyping across the Uppsala cohort. This marker was previously reported to contain CAA repeats with 10, 11 and 13 copies (Shikata *et al.*, 2004).

2.3 MASS SPECTROMETRY GENOTYPING

2.3.1 Assay design

Flanking sequences of selected SNPs were downloaded from the Ensembl database (http://www.ensembl.org). The sequences were filtered for repetitive sequences by RepeatMasker (http://repeatmasker.org), and assays were designed with MassARRAY Assay design v3.1 (Sequenom).

2.3.2 PCR amplication of SNP loci

2 μl purified DNA with a concentration of 2 ng/μl for MassEXTEND or 4 ng/μl for iPLEX were dispensed into 384-well microtiter plates using a MULTI-MEK 96 dispenser (Beckman). For genotyping quality control, each 384-well microtiter plate contained two internal duplication samples in 4 wells and water in 16 wells. Before setting up the PCR, the 384-well plate containing the DNA was briefly centrifuged.

2.3.2.1 MassEXTEND

In the MassEXTEND assay, the PCR step is variable depending upon the plexing levels. If each reaction was plexing lower than 8, i.e. detecting less than 8 SNPs in a single reaction, 1 μl PCR mix containing 0.04 μl BD Titanium Taq polymerase (BD BioSciences, Clontech), 0.75 μl PE PCR buffer, 0.2 μl 25m M dNTP and 0.01 μl deionised water was mixed with 2 μl 375 nM primer mix. The total 3 μl mix was then added into the 384-well microtiter plate, which was then centrifuged at 2000 rpm for 1 minute.

When the assay had 8-plex or more, 1 μl PCR mix containing 0.04 μl BD Titanium Taq polymerase, 0.5 μl BD Titanium PCR buffer, 0.2 μl 25mM dNTP and 0.26 μl deionised water was mixed with 2 μl 500 nM primer mix. The total 3 μl mix was then added into the 384-well microtiter plate, which was then centrifuged at 2000 rpm for 1 minute. The thermal cycling conditions were same for both low- and high-plex MassEXTEND assays. Initially the plates were heated at 95℃ for 1 minute, followed by 45 cycles of 95℃ for 20 seconds, 56℃ for 30 seconds, and 72℃ for 1 minute, and a final step at 72℃ for 3 minutes.

2.3.2.2 iPLEX

The composition of the iPLEX PCR reaction is different to that of the MassEXTEND. The PCR setup includes: 8 ng genomic DNA in 2 μl deionised water, 1μl 500 nM primer mix and 2 μl PCR mix (composed of 0.1 μl Qiagen Hotstar Taq polymerase, 0.1 μl 25mM dNTP; 0.325 μl 25mM $MgSO_4$, 0.625 μl Qiagen PCR buffer and 0.85 μl deionised water). The plate was then centrifuged at 2000 rpm for 1 minute. The thermal cycling condition is slightly different to the MassEXTEND assay because of different enzyme used. Initially the 384-well microtiter plates were heated at 94℃ for 15 minute, followed by 45 cycles of 94℃ for 20 seconds, 56℃ for 30 seconds, and 72℃ for 1 minute, and a final step at 72℃ for 3 minutes.

2.3.3 Shrimp alkaline phosphatase treatment

After PCR, shrimp alkaline phosphatase (SAP) treatment is performed to remove the 5'-phosphate group from DNA. The removal of these 5'-phosphate groups is to prevent the PCR products from ligating to each other and the excess dNTPs. 2 μl SAP mix was applied into each well of the 384-well microtiter plate. The plate was then spun at 2000 rpm for 1 minute,

followed by incubation in a thermal cycler at 37°C for 20 minutes, followed by 5 minutes at 85°C to inactivate the enzyme. When the reaction was complete, the plates were re-centrifuged at 2000 rpm for 1 minute. For MassEXTEND the 2 µl SAP mix includes 0.3 µl SAP, 0.2 µl 10x TS buffer and 1.5 µl deionised water whilst for iPLEX it includes 0.3 µl SAP, 0.17 µl 10x SAP buffer and 1.53 µl deionised water.

2.3.4 Oligo extension

Similar to the composition of MassEXTEND and iPLEX PCRs, the oligo extension reaction is different for high and low MassEXTEND, as well as the iPLEX.

2.3.4.1 MassEXTEND

For an oligo extension reaction lower than 8-plex, 2 µl extension mix was added to the 384-well PCR plate including 0.018 µl Thermosequenase, 0.2 µl 10X Thermosequenase buffer, 0.9 µl 0.5mM stop mix and 0.382 µl deionised water was mixed with 0.5 µl 10µM extension primer; for a reaction equal or higher than 8-plex, 1 µl extension mix was added to the 384-well PCR plate including 0.04 µl Thermosequenase, 0.2 µl 10X Thermosequenase buffer, 0.45 µl 1mM stop mix and 0.31 µl deionised water was mixed with 1 µl 9µM extension primer. The plates were then placed on a heated thermal cycler at 94℃ for 2 minutes, followed by 55 cycles of 94℃ for 5 seconds, 52℃ for 5 seconds, and 72℃ for 5 seconds, and a chill down step at 10℃ until the plate was removed from the machine. Importantly, the extension for high-plex assay needs 99 cycles of 94℃ for 5 seconds, 52℃ for 5 seconds, and 72℃ for 5 seconds, compared with 55 cycles for low-plex assay.

2.3.4.2 iPLEX

2 μl extension mix was added to the 384-well microtiter plate which included 0.041 μl iPLEX enzyme, 0.2 μl 10X iPLEX buffer, 0.2 μl iPLEX termination mix, 0.559 μl deionised water and 1 μl primer mix (5.5 μM for low mass extension primers and 1.1 μM for high mass ones). The plates were then placed on a heated thermal cycler at 94°C for 30 seconds, followed by 45 cycles of 94°C for 5 seconds, 52°C for 5 seconds, and 80°C for 5 seconds, and a final step at 72°C for 3 minutes. The plate was then chilled down to 10°C until removed from the machine.

2.3.5 Desalting

After oligo extension, the plate was briefly centrifuged and added with 16 μl HPLC grade water added to each well. Resin was then applied onto a 384-well dimpled plate and it was left at room temperature for 15 minutes to let the resin dry out. The dimpled resin plate was then inverted onto the 384-well microtiter plate and to allow transfer of the resin into each well. The plate was then sealed and left on the rotator for 15 minutes. The plate was subsequently centrifuged at 4000 rpm for 6 minutes before it was ready for spotting.

2.3.6 Sample spotting and analysis

Before spotting, it was important to make sure that the resin was well precipitated in the bottom of each well of the 384-well microtiter plate. The reaction products on the 384-well microtiter plates were transferred onto SpectroCHIP (Sequenom) with MassARRAY Nanodispenser (Sequenom). The SpectroCHIP was then analysed with MassARRAY

Compact Analyzer (Sequenom), and allele calling was performed using TyperAnalyzer v3.4 (Sequenom).

On each SpectroCHIP (i.e. 384 samples), the result from 8 replicated samples and 16 samples including only water were analysed for quality control. Each SNP assay with more than 4 calls from the 16 water samples will be disregarded, whilst data from any chip with more than 2 inconsistent replication results from the 8 replicated samples will be also disregarded.

2.4 OTHER GENOTYPING

2.4.1 Taqman genotyping

Detecting the CYP2C9*2 and CYP2C9*3 SNPs was performed by Taqman SNP Genotyping assay (Applied Biosystems). Pre-Developed Assay Reagents for Allelic Discrimination (TaqMan PDARs for AD) uses the 5’ nuclease assay to genotype purified DNA samples and discriminate between the two alleles of a SNP.

For discriminating between the two alleles a 5 μl reaction was setup which contained 2.5 μl 2x Taqman Universal PCR Master Mix, 0.5 μl 10x Allelic Discrimination Mix and 8 ng genomic DNA (in 2 μl deionised water). The PCR reaction was carried out on the MJ thermal cycler including the first initial step at 50°C for 2 minutes to optimise AmpErase UNG enzyme activity and 95°C for 10 minutes to activate AmpliTaq Gold DNA polymerase, and the amplification conditions were 55 cycles for 92°C for 15 seconds and 60°C for 90 seconds.

The reaction plate was read using ABI PRISM 7900HT Sequence Detection System. The genotypes were analysed with SDS software (Applied Biosystem) and the two alleles were called manually in accordance with the allele clusters.

2.4.2 Microsatellite genotyping for GGCX

Genotyping was performed by PCR amplification using HEX-fluorescent labelled primers (Table 2.1). Forward and reverse primers designed to amply the repeated region were generated using Primer 3 (http://frodo.wi.mit.edu/).

Table 2.1. Oligo sequence for amplying microsatellite in intron 6 of GGCX.

Primer	*Sequence*
GGCX-Forward	ggatatgttagaaaacattgaacacc
GGCX-Reverse	gtggctgggtagatgcctaag

PCR reaction was performed in a 25 µl reaction mix containing 2.5 µl 10x NEB buffer, 2.5 µl 2 mM dNTP, 1.6 µl primer mix (5 µM), 0.125 µl Taq polymerase, 10 ng patient DNA and 17.3 µl distilled water. Three different thermal cycling conditions were tested (see Table 2.2). The PCR products (amplicons) were electrophoretically-separated on an ABI PRISM Genetic Analyzer (Applied Biosystems).

Table 2.2. Thermo cycling conditions of GGCX microsatellite PCR.

Condistion	Step 1	Step 2 (10 cycles)			Step 3 (30 cycles)		
1	94°C	93°C	65°C	72°C	93°C	60°C	72°C
2	94°C	93°C	60°C	72°C	93°C	55°C	72°C
3	94°C	93°C	55°C	72°C	93°C	50°C	72°C
Duration	5 min	30 sec	50 sec	50 sec	30 sec	50 sec	50 sec

Allele calling was performed with the Genotyper Software v3.7 (Applied Biosystems). In the 201 Swedish warfarin-treated patients, six different genotypes were detected and the genotype and length of corresponding amplicon are listed in Table 2.3.

Table 2.3. GGCX microsatellite genotype and expected amplicon

Genotype	Repeat number	Amplicon length (bp)
1	16	436
2	15	433
3	14	430
4	13	427
5	12	424
6	11	421
7	10	418

2.5 EXON RESEQUENCING

2.5.1 Introduction

Genes which have been implicated to play a role in causing bleeding in warfarin patients were chosen for re-sequencing on the bleeders. Twenty-four bleeders from Uppsala cohort and the extra 12 bleeders, together with 12 control Swedish warfarin patients without bleeding episodes, were included in re-sequencing the candidate genes.

2.5.2 Primer design

The genes chosen for resequencing were firstly manually annotated and curated with the help from the Sanger Institute HAVANA group (http://www.sanger.ac.uk/HGP/havana/). Gene sequences, plus 1 kb flanks on both ends, were extracted from the internal sequence database and the primers amplifying each exons for each gene were automatically designed.

Without exception, 125 bp flanks on both ends of each exon were included in the region to be sequenced. Each amplicon (fragment amplified by PCR reaction) is around 500 to 550 bp to ensure perfect sequence read quality. For the exons which are too large to be amplified by 1 amplicon, multiple amplicons were adopted to toggle the exons and a minimal 100 bp overlap was forced. However, in some cases, the automatic amplicon design didn't meet the above criteria and the primers were designed manually.

Primer 3 (http://frodo.wi.mit.edu/) was used to select primers in manual process. Up to 10 amplicons were reported in each design attempt by the software. The first amplicon is

designed as described above whilst the next amplicon targets 100 bases from the 3' end of upstream amplicons. This tolerates 50 'messy' bases at the start of each read. If no amplicon could be designed, the target moves outward 50 bases and the search of primers would be attempted again - this procedure iterates until the whole exon is well included in amplification. Successfully designed primers were then compared to the golden path sequence with IPCRESS developed internally at the Sanger Institute. Any primers with 2 or more matches were removed and the amplicons were recycled.

2.5.3 PCR and sequencing

The PCR was performed using a 384-well PCR plate. 7.5 ng genomic DNA (in 7.5 μl deionised water) was used for each reaction. The 7.5 μl PCR mix (composed of 2.5 μl PCR buffer, 2.5 μl 1 mM dNTP, 0.15 μl HotStar Taq polymerase and 2.35 μl deionised water) was then added into each well.

The thermo cycling conditions included a initial heated step at 95°C for 15 minutes, followed by 39 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds and a final cycle of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 10 minutes. The PCR products were then treated with Shrimp Alkaline Phosphatase at 37°C for 30 minutes followed by 80°C for 15 minutes to inactivate the enzyme.

Sequencing experiments were performed by dedicated sequencing team at the Sanger Institute. The reaction was performed using the same primers as the PCR amplification step in both forward and reverse directions to produce double-stranded sequence. The protocol was modified in-house and each 5 μl reaction contained 1 μl diluted PCR product, 2 μl

diluted primer (15 ng/μl) and 2 μl sequencing mix (0.12 μl BigDye (v3.1), 1.38 μl Sanger BigDye reaction buffer (v2), 0.46 μl double distilled water (DDW) and 0.04 μl dGTP BigDye v3.0). The plates were centrifuged to ensure the content was in the bottom of each well and then placed on the MJ Thermo cycler. The cycling conditions were firstly at 96°C for 30 seconds followed by 45 cycles of 92°C for 8 seconds, 50°C for 8 seconds, and 60°C for 2 minutes, and then chilled down to 10°C until the plate removed. 30 μl of standard sequencing precipitation mix (770 ml ethanol, 16 ml 3 M sodium acetate, 188 ml DDW) was then added to each well and the plate was centrifuged at 4000 rpm at 4°C for 30 minutes. The reaction product was then analysed in ABI 3730 DNA Analyzer.

2.5.4 Sequence analysis

After sequencing experiments, the sequence reads are stored in an internal trace database. ‘Capminder’ has been adapted to update read length pass information for SNP, and ‘Exotrace’ was developed in-house by Steven Leonard to compare the sequence reads and to identify SNPs.

2.6 VKORC1 EXPRESSION IN HUMAN LIVER

2.6.1 RNA extraction

25 human liver biopsies were provided courtesy of Dr Ana Alfirevic (Liverpool) and were stored in liquid nitrogen prior to processing. To extract the RNA, the liver tissues were homogenised in 1ml of Trizol reagent. Homogenised tissue lysate was transferred into an RNase free 1.5 ml tube with 0.22 ml chloroform. The tube was vortexed thoroughly and incubated at room temperature for 3 minutes before being spun at 10000 rpm at 4°C for 10 minutes. The top phase liquid was transferred to a new tube containing 0.6 ml isopropanol followed by thorough vortexing and incubation at room temperature for 10 minutes and another spin at 13000 rpm at 4°C for 10 minutes. The isopropanol was removed and the precipitated pellet was washed with 1 ml of 75% ethanol followed by a centrifugation at 13000 rpm at 4°C for 10 minutes. The ethanol was removed and the tube was dried at room temperature for 10 minutes. The pellet was finally resuspended in 40 μl RNase free water. However, the electrophoresis result indicates that the RNA had been degraded during the preservation of the liver biopsy (also see Figure 4.7).

The extracted RNA was treated with DNase to remove residual DNA contamination. 50 μg RNA was aliquot and treated with 8 units of DNase I at 37°C for 40 minutes (GibcoGRL) followed by extraction with phenol/chloroform (Sigma-Aldrich). After centrifugation, the aqueous layer was transferred to a new 1.5 ml RNase free tube and added with 0.1 volume of 3 M sodium acetate and 3 volumes of 100% ethanol. The tube was incubated at -80°C for an hour followed by a centrifugation at 4°C for 15 minutes. The ethanol was removed and the pellet was washed with 100 μl of 75% ethanol followed by centrifugation at 4°C for 10

minutes. The ethanol was removed and the pellet was resuspended in 20 μl DEPC treated water and stored at -80°C.

2.6.2 Assay of VKORC1 mRNA

The DNase treated RNA samples were converted to cDNA prior to be analysed for *VKORC1* expression. The reverse transcription was carried out with Promega Reverse Transcription System containing 1 μg RNA, 4 μl $MgCl_2$ (25 mM), 2 μl Reverse Transcription 10X Buffer, 2 μl dNTP Mixutre (10 mM), 0.5 μl Recombinant RNasin Ribonuclease Inhibitor 15 units AMV Reverse Transcriptase , 0.5 μg oligo $(dT)_{15}$ primer in 20 μl reaction. The tube was incubated at 42°C for 15 minutes, 95°C for 5 minutes, 0 ^{o}C for 5 minutes and stored at -20 ^{o}C.

1.2 μl of total cDNA was used as template for the quantitative PCR in the presence of SYBR green reporter (Applied Biosystems), as described by Rieder and colleagues (Rieder et al. 2005). Each 25 μl reaction included 12.5 μl 2X SYBR Green PCR Master Mix, 3.0 μl forward and reverse primer (5 μM), 5 μl total cDNA and double distilled water and was performed on an Applied Biosystems 7900HT with standard thermo cycling conditions: 95 ^{o}C for 15 minutes, 40 cycles of 94 ^{o}C for 30 seconds, 60 ^{o}C for 30 seconds and 72 ^{o}C for 30 seconds, and 72 ^{o}C for 5 minutes. The *VKORC1* mRNA level were normalised to *GAPDH* expression levels in each liver sample. The primer sequences on *VKORC1* and *GAPDH* are listed in Table 2.4.

Table 2.4. Primer sequences for *VKORC1* and *GAPDH* gene expression.

Primer	**Sequence**
VKORC1-F	ATCAGCTGTTCGCGCGTC
VKORC1-R	AGAGCACGAAGAACAGGATC
GAPDH-F	ACAGTCAGCCGCATCTTCTT
GAPDH-R	ATGGGTGGAATCATATTGGAAC

2.7 COMPUTATIONAL ANALYSIS

2.7.1 Data processing

SNP genotyping results were stored in an internal database. To assess the data quality control, a PERL script command '~prima/bin/perl/msPlateStatL.pl' was used to produce the genotyping report to check self-priming in negative control (water) and discrepancy calls from duplicate samples. Genotype calls passing quality controls were extracted from the database using a PERL script command '~prima/bin/perl/genoMspec1PlateCallAgg.pl'. Finally, the genotype call file was formatted using a PERL script command 'CallFileFormat.pl', and all the residual discrepant calls were removed. The scripts are listed in Table 2.5.

Table 2.5 Scripts used for data processing.

Script command	Function	Programmed
~prima/bin/perl/msPlateStatL.pl	Data quality check (self-priming and discrepant call)	Jilur Ghori
'~prima/bin/perl/genoMspec1PlateCallAgg.pl	Extract genotyping result from database, including aggressive calls	Jilur Ghori
CallFileFormat.pl	Format the genotyping call file to expected format.	Leslie Chen

2.7.2 Genotype analysis

Genotyped SNPs were analysed and visualised by Haploview (Barrett et al. 2005) which was downloaded from http://www.broad.mit.edu/mpg/haploview/. Analyses including individual patient or SNP exclusion, LD and haplotype block analysis and visualisation, Hardy-Weinberg Equilibrium (HWE) test, and tag SNPs selection were performed in Haploview.

Linkage format files (PED files) recording SNP genotype information were generated with a PERL script described earlier. SNP identifiers and positions were specified in a separate file (INFO file). Individual patients with 50% missing genotype SNPs were excluded for analyses. The haplotype blocks were identified using a predefined definition according to Gabriel and colleagues (Gabriel et al. 2002). The SNP exclusion list was defined with the criteria listed in Table 2.6.

Table 2.6. SNP exclusion criteria in Haploview.

Criteria	**Threshold**
HWE p-value cutoff	0.0010
Minimal genotype call rate	0.7
Maximum number of medel error	1
Minimum MAF	0.05

SNPs giving the same information were excluded in the WARG study, and tag SNPs were selected with the implementation of Tagger (de Bakker et al. 2005). Three different tagging methods can be applied including pairwise tagging with specified r^2 threshold and aggressive tagging using either 2-marker or 2- and 3-marker haplotypes with specified LOD threshold. The tag SNP selection applied in this thesis used a specified r^2 threshold of 0.8.

2.7.3 Multiple sequence alignment

Multiple sequence alignment analysis for both DNA and protein sequence was performed with ClustalX v1.83 (Jeanmougin et al. 1998) downloaded from European Bioinformatics Institute (EBI, ftp.ebi.ac.uk). The DNA and protein sequences were downloaded from

National Center for Biotechnology Information, U.S. National Library of Medicine (http://www.ncbi.nlm.nih.gov/).

2.7.4 Evolutional conserver region (ECR) analysis

To identify potential regulatory regions of *VKORC1*, a comparison for ECR was performed using genomic sequences from human, chimpanzee, mouse, rat, dog and chicken. Analysis was performed using web-based programme, zPicture (Ovcharenko et al. 2004). Human *VKORC1* genomic sequence was used to blast genomes of chimpanzee, mouse, rat, dog, and chicken with Ensembl genome browser (http://www.ensembl.org). The corresponding genomic sequences of the other species were uploaded to the multi-zPicture website (http://zpicture.dcode.org/multiz.php) with human genomic sequencing containing *VKORC1* region.

2.8 STATISTICAL ANALYSIS

2.8.1 Statistics in Uppsala study

Univariate and multivariate regression analysis was assisted by Niclas Eriksson (Uppsala). SAS (SAS Institute Inc) and SPLUS (Insightful Corporation) software were used to perform both univariate and multiple analyses of predictor's impact on the square root of warfarin dose using linear regression models. To address the partial dependence among tests of SNPs in LD, Bonferroni correction for multiple testing based on calculating the effective number of independent tests (Meff) was tested and a Permutation procedure was also applied (Cheverud 2001; Li 2001; Nyholt 2004).

QTPhase component of Unphased software (Dudbridge 2003) was used to estimate haplotype frequencies for *VKORC1* genotypes. Unphased is developed by Frank Dudbridge at MRC Biostatistics Unit, University of Cambridge, United Kingdom and was downloaded from http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/. The means and variances of inter-individual dose were calculated in association with each haplotype. Differences among the haplotype means were also statistically tested with QTPhase.

2.8.2 Statistics in WARG study

Univariate and multivariate regression analysis was assisted by Niclas Eriksson (Uppsala). R version 2.5.1 (http://www.r-project.org/, R foundation for statistical computing,) and SAS version 9.1.3 (SAS Institute Inc) were used for statistical analyses in the WARG study. Univariate and multivariable analyses of predictor impact on the square root of warfarin dose

were calculated using linear regression analyses. To account for partial dependence among tests of SNPs in LD, Bonferroni correction for multiple testing was applied based on the effective number of independent tests (Meff) calculated by a spectral decomposition method (Cheverud 2001; Li 2001; Nyholt 2004).

The coefficient of determination, R^2, was used to measure the proportion of explained variance. Association with over-anticoagulation was evaluated with Log Rank test. Hazard ratios were estimated with Cox regression analyses. The prediction models were based on verified findings and only nominally significant variables ($p<0.05$) were allowed in the final mode, and the R^2 values were calculated.

2.8.3 Statistics in Case/control association

The case/control analysis of warfarin bleeders was performed with PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) software developed by Shaun Purcell at the Center for Human Genetic Research, Massachusetts General Hospital, and the Broad Institute of Harvard and MIT.

A specified linkage format file (PED file) is needed and contains the individual patient's phenotype and genotype information. SNP marker identifiers and positions were recorded in a separate file (MAP file). Various case-control tests including Cochran-Armitage trend test, Fisher's exact test, two- and three-marker haplotype test using sliding window, recessive gene action test, and gene-gene interaction tests were performed for bleeding complication association. For the cell number smaller than 5 in the recessive gene action test, the p-value, odds ratio and 95% confident interval were calculated with R version 2.6.0.