Dipartimento di Medicina e Chirurgia
Corso di Dottorato in Medicina Traslazionale e Clinica
XIV Ciclo

TESI DI DOTTORATO

*Translational pharmacogenomics: a study of warfarin dosing*

Coordinatore
Ch.ma Prof. Amelia Filippelli

Candidata
Giusy Russomanno

Tutor
Ch.ma Prof. Viviana Izzo

Tutor Esterno
Ch.mo Prof. Sir Munir Pirmohamed

Anno Accademico 2015/2016
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>i</td>
</tr>
<tr>
<td>SOMMARIO</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>iv</td>
</tr>
<tr>
<td>CHAPTER 1: General Introduction</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2: Pharmacogenetic-based warfarin dosing algorithm in Southern Italians</td>
<td>45</td>
</tr>
<tr>
<td>CHAPTER 3: Circulating miRNAs as novel biomarkers of response to warfarin</td>
<td>90</td>
</tr>
<tr>
<td>CHAPTER 4: Warfarin Extreme Phenotypes: A Genome-Wide Association Study</td>
<td>141</td>
</tr>
<tr>
<td>CHAPTER 5: Final Discussion</td>
<td>188</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>204</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>232</td>
</tr>
</tbody>
</table>
ABSTRACT

Warfarin is one of the most commonly used oral anticoagulants worldwide and is highly efficacious for the treatment and prevention of thromboembolic disorders. However, due to its narrow therapeutic index, large inter-individual variability in dose requirements, and extensive drug and food interactions, warfarin remains a challenging drug to prescribe. Genetic factors (CYP2C9 and VKORC1), together with clinical factors (age and body weight), account for up to 60% of warfarin dose variance, whereas ~40% variability remains poorly understood.

Several warfarin dosing algorithms, comprising genetic and non-genetic covariates have been published over the years. However, none of the published algorithms included patients from Southern Italy. We therefore conducted a candidate-gene study to develop an algorithm for predicting warfarin maintenance dose in patients from the Campania Region (n=266) in Southern Italy. Our pharmacogenetic dosing algorithm consisted of six variables (age, body surface area, amiodarone intake, CYP2C9*2, CYP2C9*3, and VKORC1 -1639G>A). It led to the accurate prediction of warfarin maintenance dose in 44% of patients (mean absolute error 7.41 mg/week). The prediction accuracy of the pharmacogenetic algorithm was superior to three previously published pharmacogenetic algorithms derived from patients in Northern and Central Italy.

Given that previous studies suggested a role for miR-133a in warfarin response, we conducted a pilot study comparing baseline serum levels of miRNA in patients who achieved warfarin stable dose (n=10) to those who did not achieve warfarin stability (n=10), using the Affymetrix miRNA array. No association was found between miRNA-133a and warfarin response. Interestingly, circulating levels of miR-548a-3p were observed to be higher in patients who did not achieve warfarin stability (P=0.0053, fold change =1.66) compared to patients who achieved stable dose. In silico analyses showed that several target genes of miR-548a-3p are involved in the coagulation pathway. Work is currently underway to validate and replicate these findings in a larger cohort of prospectively recruited patients initiated onto warfarin therapy (n=980) using TaqMan miRNA real-time quantitative PCR.

Pharmacogenetic algorithms have shown that common variants in CYP2C9 and VKORC1 genes cannot fully explain the extreme dose requirements in individuals sensitive and resistant to warfarin. To investigate the role of other genetic variants in these patients with extreme phenotypes, we performed a genome-wide association study (GWAS) comprising of warfarin sensitive patients (≤1.5 mg/day, n=55), warfarin resistant patients (≥10 mg/day, n=51), and healthy controls from the National Blood Service (NBS, n=2,501). Our results suggested that an intergenic variant on chromosome 10, rs4918797, could be involved in warfarin sensitivity. Intronic SNPs in MIR6873 on chromosome 6 (rs114213056) and PIGN on chromosome 18 (rs10163900, rs76455916, rs77118150, and rs79434376) showed suggestive association with warfarin resistance.

The findings of this thesis showed that a multitude of factors affect warfarin dosing, some of which still need further investigations. Insights of the roles of other factors such as non-coding RNA and rare genetic variants will hopefully improve dose prediction and drug efficacy and ultimately patient outcomes. The work being undertaken with warfarin acts as a pathfinder project, the concepts from which could be applied to other drugs with variable dose requirement.
Il warfarin è un farmaco anticoagulante orale largamente utilizzato per la prevenzione e il trattamento di eventi tromboembolici. Il trattamento con warfarin, tuttavia, risulta poco maneggevole a causa della sua stretta finestra terapeutica, dell’elevata variabilità inter-individuale nella dose richiesta e delle numerose interazioni con farmaci e alimenti. Fattori genetici (CYP2C9 e VKORC1) e variabili cliniche (età e peso) spiegano circa il 60% della variabilità nella dose che si osserva nei pazienti in trattamento, mentre il 40% circa è legato a fattori non ancora noti.

Diversi algoritmi per determinare la dose individuale di warfarin sulla base di variabili cliniche e genetiche sono stati sviluppati nel corso degli anni, ma nessuno degli studi italiani pubblicati finora ha incluso pazienti del Sud Italia. Per tale ragione, abbiamo condotto uno studio geno-candidato al fine di sviluppare un algoritmo farmacogenetico per calcolare la dose di mantenimento di warfarin in 266 pazienti Campani. Un algoritmo a sei variabili (età, superficie corporea, uso di amiodarone, CYP2C9*2, CYP2C9*3 e VKORC1 -1639G>A) è stato derivato, permettendo l’accurata determinazione della dose nel 44% dei pazienti considerati (errore medio assoluto di 7.41 mg settimanali). Il potere predittivo di tale algoritmo farmacogenetico si è dimostrato superiore a quello di tre algoritmi farmacogenetici pubblicati includendo pazienti del Centro e Nord Italia.

Sulla base di studi che suggerivano un possibile ruolo di miR-133a nel meccanismo di risposta al warfarin, abbiamo condotto uno studio pilota per confrontare i livelli basali di espressione dei microRNA (miRNA) (Affymetrix miRNA array) in pazienti in trattamento con warfarin che raggiungevano (n=10) o meno (n=10) la dose stabile. Mentre per miR-133a non è stata trovata nessuna differenza significativa, i livelli sierici di miR-548a-3p erano più elevati nei pazienti che non raggiungevano la dose stabile (P=0.0053). Analisi in silico hanno dimostrato che diversi geni target di miR-548a-3p sono coinvolti nella cascata della coagulazione. I risultati ottenuti sono al momento in fase di validazione e replicauzione in una coorte prospettica di pazienti in trattamento con warfarin (n=980) mediante TaqMan miRNA real-time quantitative PCR.

Considerato che polimorfismi genetici in VKORC1 e CYP2C9 non sono i soli determinanti di dosi estremamente basse o elevate in pazienti sensibili o resistenti al warfarin, per l’individuazione di nuove varianti genetiche coinvolte nel fenotipo estremo di risposta al farmaco, abbiamo condotto uno studio di associazione genome-wide includendo pazienti sensibili al warfarin (≤1,5 mg/die, n=55), pazienti resistenti (≥10 mg/die, n=51) e controlli sani (n=2501). Una variante intergenica nel cromosoma 10, rs4918797, è stata significativamente associata alla sensibilità al warfarin, mentre varianti introniche nei geni MIR6873 nel cromosoma 6 (rs114213056) e PIGN nel cromosoma 18 (rs10163900, rs76455916, rs77118150, e rs79434376) erano indicativi di resistenza.

I risultati di questo lavoro di tesi dimostrano che una serie di fattori influenzano la dose di warfarin. Ulteriori studi sul ruolo dei miRNA e di varianti genetiche rare potrebbero dimostrarsi utili per migliorare l’accuratezza degli algoritmi per il calcolo della dose di warfarin e, di conseguenza, aumentare la sicurezza e l’efficacia del trattamento. Infine, l’approccio utilizzato in questo studio potrebbe essere applicato anche ad altri farmaci per i quali il dosaggio dovrebbe essere individualizzato.
ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to Professor Amelia Filippelli for her support throughout the years. Her kindness, encouragement, guidance, unsurpassed knowledge and inspiring ideas have made my study immensely enjoyable and rewarding. I would not have come this far without her help. Thanks go out to my supervisor Professor Viviana Izzo, for being such a fabulous mentor, guide and great friend.

A big thank to Professor Sir Munir Pirmohamed for giving me this invaluable opportunity to study and research at the Wolfson Centre for Personalised Medicine in Liverpool. It has been an honour to work with you. Thanks for all the inspiring and motivating discussions, you have taught me many lessons in life that I will never forget.

This PhD thesis would have been not possible without the enormous help and kind guidance of Dr Eunice Zhang, enthusiastic researcher and lovely friend. You made me fell home since my first day in Liverpool. Thanks for sharing with me much more than just knowledge. I owe you sincere gratitude, and words will never be enough.

Thanks to Professor Lucia Sacchetti and Dr Cristina Mazzaccara for their help with the pharmacogenetic-based warfarin dosing algorithm in Southern Italians.

Thanks to Dr Ana Alfirevic for her encouraging words and advices, and to Sam and Angela for their kindness and patience with all my paper-work. Thanks for being such lovely people!

Thanks to my colleagues and friends, Valentina, Vittorio, Federica, Francesca, Ilaria, and Daniela for all the years we spent together, and to Michela, you will always be with us.

To the members of The Wolfson Centre for Personalised Medicine. You have all made my stay in Liverpool unforgettable. My special thanks to Alex, Ana, Andrea, Athina, Benedicta, Carmen, Christine, Dan, Fabio, George, Gospel, Gurpreed, James, Joe, John, Kate, Lorna, Nathalia, Philippe, Reyad, Sarah, Sam, Sammy, Stephanie, Sudeep, and to my favourite technicians, Laurence and Terry.

Words of thanks will never be enough for my best friend Giusy who has been always there for me. “We were but stones, your light made us stars”. Thanks to Mario for being such a lovely (and jealous) brother. To Ale and Carmen, because home is where a hug is (and maybe very close to Eric’s).

I owe deep gratitude towards my family especially my beloved mum and dad for their constant support, patience, encouragement and endless love. My love and thanks also goes to Myry and Francesco, who make me proud every day and give me strength and courage to go on.

Above all, to my angel Marcy.
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCB1</td>
<td>ATP binding cassette, subfamily B, member 1</td>
</tr>
<tr>
<td>ADRs</td>
<td>adverse drug reactions</td>
</tr>
<tr>
<td>AF</td>
<td>atrial fibrillation</td>
</tr>
<tr>
<td>AIFA</td>
<td>Italian Medication Agency</td>
</tr>
<tr>
<td>AMI</td>
<td>acute myocardial infarction</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>APC</td>
<td>activated protein C</td>
</tr>
<tr>
<td>APOE</td>
<td>apolipoprotein E</td>
</tr>
<tr>
<td>B.C.</td>
<td>before Christ</td>
</tr>
<tr>
<td>BMI</td>
<td>body mass index</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>body surface area</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>calcium</td>
</tr>
<tr>
<td>CALU</td>
<td>calumenin</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CEU</td>
<td>Utah residents with ancestry from northern and western Europe</td>
</tr>
<tr>
<td>CHB</td>
<td>Han Chinese in Beijing, China</td>
</tr>
<tr>
<td>Chr</td>
<td>chromosome</td>
</tr>
<tr>
<td>CM</td>
<td>concomitant medications</td>
</tr>
<tr>
<td>CO₂</td>
<td>carbon dioxide</td>
</tr>
<tr>
<td>COAG</td>
<td>Clarification of Optimal Anticoagulation through Genetics</td>
</tr>
<tr>
<td>CPB2</td>
<td>carboxypeptidase B2</td>
</tr>
<tr>
<td>CPIC</td>
<td>Clinical Pharmacogenetics Implementation Consortium</td>
</tr>
<tr>
<td>Ct</td>
<td>cycle threshold</td>
</tr>
<tr>
<td>CV</td>
<td>coefficient of variation</td>
</tr>
<tr>
<td>CVA</td>
<td>cardiovascular accident</td>
</tr>
<tr>
<td>CYP</td>
<td>cytochrome P450</td>
</tr>
<tr>
<td>CYP1A1</td>
<td>cytochrome P450, family 1, subfamily A, polypeptide 1</td>
</tr>
<tr>
<td>CYP1A2</td>
<td>cytochrome P450, family 1, subfamily A, polypeptide 2</td>
</tr>
<tr>
<td>CYP1A6</td>
<td>cytochrome P450, family 1, subfamily A, polypeptide 6</td>
</tr>
<tr>
<td>CYP2A6</td>
<td>cytochrome P450, family 2, subfamily A, polypeptide 6</td>
</tr>
<tr>
<td>CYP2C115P</td>
<td>cytochrome P450, family 2, subfamily C, polypeptide 115, pseudogene</td>
</tr>
<tr>
<td>CYP2C18</td>
<td>cytochrome P450, family 2, subfamily C, polypeptide 18</td>
</tr>
<tr>
<td>CYP2C19</td>
<td>cytochrome P450, family 2, subfamily C, polypeptide 19</td>
</tr>
<tr>
<td>CYP2C59P</td>
<td>cytochrome P450, family 2, subfamily C, polypeptide 59, pseudogene</td>
</tr>
<tr>
<td>CYP2C8</td>
<td>cytochrome P450, family 2, subfamily C, polypeptide 8</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>cytochrome P450, family 2, subfamily C, polypeptide 9</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>cytochrome P450, family 2, subfamily D, polypeptide 6</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>cytochrome P450, family 3, subfamily A, polypeptide 4</td>
</tr>
<tr>
<td>CYP3A5</td>
<td>cytochrome P450, family 3, subfamily A, polypeptide 5</td>
</tr>
<tr>
<td>CYP4F2</td>
<td>cytochrome P450, family 4, subfamily F, polypeptide 2</td>
</tr>
<tr>
<td>Term</td>
<td>Definition/Description</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DAVID</td>
<td>Database for Annotation, Visualization and Integrated Discovery</td>
</tr>
<tr>
<td>DHPLC</td>
<td>Denaturing high performance liquid chromatography</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>E</td>
<td>Efficiency</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>ENCODE</td>
<td>Encyclopedia of DNA Elements</td>
</tr>
<tr>
<td>EPHX1</td>
<td>Epoxide hydrolase 1</td>
</tr>
<tr>
<td>EU-PACT</td>
<td>European Pharmacogenetics of Anticoagulant Therapy</td>
</tr>
<tr>
<td>F2</td>
<td>Coagulation factor II</td>
</tr>
<tr>
<td>F5</td>
<td>Coagulation factor V</td>
</tr>
<tr>
<td>F7</td>
<td>Coagulation factor VII</td>
</tr>
<tr>
<td>F9</td>
<td>Coagulation factor IX</td>
</tr>
<tr>
<td>F10</td>
<td>Coagulation factor X</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>FG</td>
<td>Fibrinogen</td>
</tr>
<tr>
<td>GAS6</td>
<td>Growth-arrest-specific protein 6</td>
</tr>
<tr>
<td>GGCX</td>
<td>γ-Glutamyl carboxylase</td>
</tr>
<tr>
<td>Gla</td>
<td>γ-Carboxyglutamic acid</td>
</tr>
<tr>
<td>Glu</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association study</td>
</tr>
<tr>
<td>H₂O</td>
<td>Water</td>
</tr>
<tr>
<td>HDL</td>
<td>High-density lipoprotein</td>
</tr>
<tr>
<td>HGP</td>
<td>Human Genome Project</td>
</tr>
<tr>
<td>HMG-CoA</td>
<td>3-Hydroxy-3-methylglutaryl-coenzyme A</td>
</tr>
<tr>
<td>HMGIC</td>
<td>High-mobility group protein Isoform I-C</td>
</tr>
<tr>
<td>HMWK</td>
<td>High molecular weight kininogen</td>
</tr>
<tr>
<td>HSD3B7</td>
<td>Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 7</td>
</tr>
<tr>
<td>HWE</td>
<td>Hardy-Weinberg Equilibrium</td>
</tr>
<tr>
<td>IBD</td>
<td>Identity by descent</td>
</tr>
<tr>
<td>INR</td>
<td>International normalized ratio</td>
</tr>
<tr>
<td>ISI</td>
<td>International sensitivity index</td>
</tr>
<tr>
<td>IWPC</td>
<td>International Warfarin Pharmacogenetics Consortium</td>
</tr>
<tr>
<td>JPT</td>
<td>Japanese in Tokyo, Japan</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage disequilibrium</td>
</tr>
<tr>
<td>LHFPL3</td>
<td>Lipoma high-mobility group protein Isoform I-C fusion partner-like 3</td>
</tr>
<tr>
<td>LOC101928762</td>
<td>RNA gene, affiliated with the ncRNA class</td>
</tr>
<tr>
<td>LRT</td>
<td>Likelihood ratio test</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean absolute error</td>
</tr>
</tbody>
</table>
MAF  minor allele frequency
MASP1/2  mannan-binding lectin serine peptidase 1 and 2
MDR1  multidrug resistance protein 1
MeV  MultiExperiment Viewer
MGB  minor groove binder
MgCl$_2$  magnesium chloride
MI  myocardial infarction
miRNA  microRNA
mRNA  messenger RNA

NAD(P)H  nicotinamide adenine dinucleotide phosphate
NBS  National Blood Service
NCBI  National Centre for Biotechnology Information
ncRNA  non-coding RNA
NFQ  non-fluorescent quencher
(NH$_4$)$_2$SO$_4$  ammonium sulphate
NOAC  novel oral anticoagulant
NQO1  NAD(P)H dehydrogenase, quinone 1
NR1I2  nuclear receptor subfamily 1, group I, member 2
NR1I3  nuclear receptor subfamily 1, group I, member 3
NSAIDs  nonsteroidal anti-inflammatory drugs

O$_2$  oxygen
ORM1  orosomucoid 1
ORM2  orosomucoid 2

PCA  principal-component analysis
PCR  polymerase chain reaction
PE  pulmonary embolism
PIGN  phosphatidylinositol glycan anchor biosynthesis, class N
PK  prekallikrein
PLG  plasminogen
PROC  protein C
PROS1  protein S
PROZ  protein Z
PT  prothrombin time

Q  quencher
QC  quality control
qPCR  quantitative PCR

RCT  randomised controlled trial
RT  reverse transcription

SD  standard deviation
SE  standard error
SERPINC1  serpin peptidase inhibitor, clade C (antithrombin), member 1
SETD1A: SET domain containing 1A
SNORA: small nucleolar RNA, H/ACA box
SNORD: small nucleolar RNA, C/D box
SNP: single nucleotide polymorphism
SPSS: Statistical Package for the Social Sciences
Std: standard
STX1B: syntaxin 1B
TAC: Transcriptome Analysis Console
TBE: tris borate EDTA
TE: Tris-EDTA buffer
TEAA: triethylammonium acetate
TF: tissue factor
TFPI: tissue factor pathway inhibitor
TIA: transient ischaemic attack
Tris-HCl: 2-amino-2-hydroxymethyl-1,3-propanediol hydrochloride
TTR: time in the therapeutic INR range
UTR: 3’ untranslated region
UV: ultraviolet
VDR: vitamin D receptor
VKOR: vitamin K 2,3-epoxide reductase
VKORC1: vitamin K epoxide reductase complex, subunit 1
VTE: venous thromboembolism
vWF: von Willebrand factor
WEP: warfarin extreme phenotype study samples
YRI: Yoruba in Ibadan, Nigeria
CHAPTER 1

General Introduction
CONTENTS

1.1 Variability in drug response ................................................................. 3
1.2 Single nucleotide polymorphism and personalised medicine... 7
1.3 Pharmacogenetics and pharmacogenomics ................................. 9
1.4 The blood coagulation system ......................................................... 11
  1.4.1 Extrinsic pathway ........................................................................ 14
  1.4.2 Intrinsic pathway ........................................................................ 14
  1.4.3 Common pathway .......................................................................... 15
  1.4.4 Coagulation pathway regulators .................................................. 16
1.5 Warfarin .............................................................................................. 16
  1.5.1 Clinical use of warfarin ................................................................. 17
  1.5.2 Pharmacokinetics of warfarin ....................................................... 20
  1.5.3 Pharmacodynamics of warfarin .................................................... 21
1.6 Clinical and environmental factors affecting warfarin response ......................................................................................... 24
1.7 Genetic contributions to warfarin response ................................. 27
  1.7.1 CYP2C9 .......................................................................................... 27
  1.7.2 VKORC1 ........................................................................................ 29
  1.7.3 CYP4F2 ........................................................................................ 31
  1.7.4 Other genetic factors .................................................................... 32
1.8 Warfarin dosing algorithms .............................................................. 38
1.9 miRNAs as potential biomarkers of therapy response .......... 41
1.10 Thesis aims ......................................................................................... 43
1.1 Variability in drug response

“Every individual is different from another and hence should be considered as a different entity. As many variations are there in the universe, all are seen in human beings”.

*Charaka Samhita*

Patients vary widely in their responses to drugs. A treatment with the “one-size-fits-all” approach that inherently ignores the marked inter-individual variation which exists in response to medications has nowadays been replaced by the idea that therapy should be personalised. *Personalised medicine* is defined as medical care for each patient’s unique condition, and its ultimate goal is to administer “the right dose of the right drug to the right person at the right time”.

The significance of personalised medicine has long been understood by clinicians since 1500 B.C. (Chatterjee and Pancholi, 2011). Ayurveda, the ancient Indian, traditional Chinese and Korean science of medicine, has long considered and practiced individualized treatment schedules (Joshi et al., 2010). Hippocrates, the father of western medicine, advocated personalized medicine. He considered factors like age, physique and patient's constitution while prescribing drugs to his patients (Sykiotis et al., 2005) since they did not respond to drug therapies in a uniform and predictable manner (Roden et al., 2011).

It is well known that drug response is a gene-by-environment phenotype. There are multiple contributory factors in drug response variation such as gender, age, body mass, diet, the presence of other drugs or of
particular disease states and exposure to certain chemicals or toxins, such as cigarette smoke (Figure 1.1) (Vesell, 1991). The individual’s response to a drug depends on the complex interplay between environmental and genetic factors.

Individual variations can influence the rate of absorption, distribution, metabolism, and elimination of the medication leading to varied plasma concentrations or excretion profiles, resulting in lack of efficacy or induced toxicity (Evans and Relling, 1999).

The inter-patient heterogeneity in drug efficacy and toxicity has a potential impact on the quality and cost of healthcare. This variability is seen, not only in beneficial responses, but also in adverse drug reactions (ADRs), an increasingly recognized problem that extracts a huge toll on lives and in healthcare costs (Brewer and Colditz, 1999; Spear et al., 2001).

Understanding the role of genetics in drug responses could thus ensure better therapeutic efficiency and reduce the incidence of ADRs (Figure 1.2). A combinatory approach evaluating clinical, environmental and genetic factors is crucial to succeed in personalising and stratifying drug therapy.
Figure 1.1. Exogenous and endogenous factors contribute to variation in drug response. The circular design shows the multiplicity of either well-established or suspected host factors that may influence drug response. The outer circular line indicates the close inter-relationship between all such factors. Arrow from each factor indicates the effect of each host factor on drug response. Adapted from Vesell, 1991.
Figure 1.2. Personalised medicine. Personalised medicine promises to transform the delivery of healthcare to patients. Its aim is to evolve from a reactive “one-size-fits-all” system towards a system of predictive, preventive, and precision care. This picture depicts how personalised medicine could classify people into smaller subsets on the basis of the therapy response from one large disease group. Genetic tests can help stratify patients to those who would respond effectively to a specific drug and those who would experience an ADR.
1.2 Single nucleotide polymorphism and personalised medicine

Each individual inherits their genome from the parents and the inherited differences in DNA sequence contribute to phenotypic variation influencing individual's anthropometric characteristics, risk of diseases and response to environment (Ye, 2005). The recognition of these inter-individual differentiations that brings variation in drug response is an essential step towards personalized medicine.

Following the conclusion of the Human Genome Project (HGP) in 2003 (International Human Genome Sequencing Consortium, 2004), substantial progress has been made in human genetics and genomics research over the past decade (Lander et al., 2001; Venter et al., 2001). Several large-scale international projects have been launched, such as the International HapMap Project, the Encyclopedia of DNA Elements (ENCODE) Project, the 1000 Genomes Project, and the 10K Genome Project (1000 Genomes Project Consortium et al., 2010, 2012, 2015; ENCODE Project Consortium, 2004; International HapMap 3 Consortium et al., 2010; International HapMap Consortium, 2003, 2005; International HapMap Consortium et al., 2007; Koepfli et al., 2015), contributing to our understanding and knowledge of human genetics and genomics.

There are over 3 billion base pairs across the human genome, encoding over 22,000 genes. Although humans are 99% identical, the 1% difference between two individuals’ genome includes more than 12 million potential variations (1000 Genomes Project Consortium et al., 2010; International HapMap Consortium et al., 2007). Genetic variants that occur
infrequently at less than 1% in the population are called “mutations”, whereas variants that occur more frequently at 1% or greater are called “polymorphisms”, and they include insertions or deletions, copy-number variations, variable number of tandem repeats and single nucleotide polymorphisms (SNPs).

Single nucleotide polymorphisms are the major source of heterogeneity, accounting for nearly 90% of all sequence variations (Collins et al., 1998). SNPs occur on average about every 100 to 300 bases along the human genome, and are present in both coding and noncoding regions. SNPs located in the coding regions are classified into (i) synonymous, and (ii) nonsynonymous. Synonymous SNPs do not affect the protein sequence, while nonsynonymous SNPs cause a change in the amino acid sequence of the protein (Hunt et al., 2009). These polymorphisms in any one gene, including those coding for drug receptors, drug transporters and cell signalling pathways can be an important determinant of clinical response and susceptibility to diseases (Nebert et al., 2013).

SNPs are evolutionarily stable, not changing much from generation to generation. The natural process of non-random association of two or more alleles at two or more physically proximate loci is known as “linkage disequilibrium” (LD). Alleles occurring together on the same chromosome and which tend to be inherited together are known as a "haplotype" (Daly et al., 2001; Gabriel et al., 2002; Reich et al., 2001). The regions of DNA that are in LD remain unchanged during recombination and thus “travel together” in transmitting the genomic material from parent to offspring. This efficiently permits the investigation of only one representative SNP that can serve as a
“tag” for nearby SNPs and haplotype blocks by decreasing the number of total SNPs that need testing directly. This has been utilised in the design of genome-wide association studies (GWAS) (Barrett and Cardon, 2006; Eberle et al., 2007; International HapMap Consortium et al., 2007; Li et al., 2008).

1.3 Pharmacogenetics and pharmacogenomics

The history of pharmacogenetics dates back to 510 B.C. when Pythagoras noted that ingestion of fava beans resulted in a potentially fatal reaction in some, but not all, individuals (Nebert, 1999). The notion that genetic variants might modulate variability in drug action was first proposed by the English physiologist Archibald Garrod (1923). He suggested that enzymatic defects not only lead to aggregation of endogenous substrates in “inborn errors of metabolism”, but also to aggregation of exogenously administered substrates, such as food, toxins and drugs, with clinical concerns (Garrod, 1923).

The concept of familial clustering of unusual drug responses was strengthened during the 1940s with the observation of a high incidence of haemolysis on exposure to antimalarial drugs among individuals with glucose-6-phosphate dehydrogenase deficiency (Beutler et al., 1955). In the 1950s, Price-Evans and colleagues identified N-acetylation as a major route of isoniazid elimination (Evans et al., 1960). Although individuals varied substantially in the extent to which a single dose of the drug was acetylated, variability between monozygotic twins was found to be small when compared with dizygotic twins (Bönicke and Lisboa, 1957), laying the groundwork for
studies that defined the clinical consequences and genetic basis of the fast and slow acetylator phenotypes (Evans, 1984; Wolkenstein et al., 1995). The past half century has seen developments in the understanding of the molecular basis of drug disposition and drug action, and of the mechanisms that determine the observed variability in drug action. Hence, the concept of a familial component in drug action initiated the field of pharmacogenetics, even before the discovery of DNA as the repository of genetic information.

The word pharmacogenetics was first used in 1959 by the German Pharmacologist Friedrich Vogel, and two years before, Arno Motulsky wrote a paper on how “drug reactions may be considered pertinent models for demonstrating the interaction of heredity and environment in the pathogenesis of disease” (Motulsky, 1957; Vogel, 1959).

Pharmacogenetics can thus be defined as the study of variability in drug response due to heredity (Nebert, 1999). Pharmacogenetics is aimed to determine the genetic differences in metabolic pathways which can affect individual responses to drugs, both therapeutically and adversely (Motulsky and Qi, 2006). The term has been coined together from the words pharmacology (the study of action of drugs in the human body) and genetics (the study of inheritance of traits). In pharmacogenetics, the analysis of a specific gene, or group of genes, may be used to predict responses to a specific drug or class of drugs. Phenotypes are scrutinized by low or exaggerated pharmacological effects, frequency of side effects, and difference in the rate of metabolism.

With the increased understanding of the molecular, cellular and genetic determinants of drug action, the emergence of the Human Genome
project, and the development of the genome sciences in the 1990s (1000 Genomes Project Consortium et al., 2010), has come the appreciation that variants in many genes might contribute to variability in drug action. Thus, the concept of using whole-genome information to predict drug action is one definition of the more recent term, pharmacogenomics (Evans and Relling, 1999; Meyer, 2000; Roses, 2000), introduced for the first time in 1997 by Marshall (Marshall, 1997). While the term pharmacogenetics is largely used in relation to genes determining drug metabolism, the latter term, pharmacogenomics, is a broader based term that encompasses all genes in the genome that may determine drug response.

Although there are constant debates in the literature as to which term should be used, both refer to the need to personalise or stratify drug therapy in order to maximise the efficacy and minimise the toxicity of drugs. For this reason, both terms are often used interchangeably (Grant, 2001; Pirmohamed, 2001).

1.4 The blood coagulation system

Coagulation is the process by which blood changes from a liquid to a gel, forming a clot. It potentially results in haemostasis, the cessation of blood loss from a damaged vessel, followed by repair. The mechanism of coagulation involves activation, adhesion, and aggregation of platelets along with deposition and maturation of fibrin.

Coagulation begins almost instantly after an injury to the blood vessel, when the endothelium lining the vessel is damaged. Disruption of the
endothelium exposes platelets to collagen in the vessel wall and clotting factor VII (FVII) to extravascular tissue factor (TF), which ultimately leads to fibrin formation. Platelets immediately form a plug at the site of injury (primary haemostasis). Other proteins, such as von Willebrand factor (vWF), facilitate the binding of platelets to the injured vessel wall. Secondary haemostasis occurs simultaneously: additional clotting factors beyond FVII respond in a complex cascade to form fibrin strands, which strengthen the platelet plug (Furie and Furie, 2005).

The coagulation cascade of secondary haemostasis has two initial pathways that lead to fibrin formation. These are the extrinsic pathway (also known as tissue factor pathway), and the intrinsic pathway (also known as contact activation pathway). The extrinsic and the intrinsic pathways both activate the final common pathway of factor X, thrombin and fibrin (Figure 1.3) (Davie and Ratnoff, 1964; Macfarlane, 1964).
Figure 1.3. The coagulation cascade. The extrinsic pathway is initiated upon vascular injury which leads to exposure of the tissue factor (TF). The contact activation pathway begins with formation of the primary complex on collagen consisting of high-molecular-weight kininogen (HMWK), prekallikrein (PK), and factor XII. The common point in both pathways is the activation of factor X to factor Xa. Factor Xa converts prothrombin to thrombin, which activates factor VIII, protein C, and factor V. Activated protein C (APC) inactivates factors Va and VIIIa. Protein Z inhibits factor Xa. Tissue factor pathway inhibitor (TFPI) inhibits the action of TF. Warfarin inhibits biologic activation of factors II, VII, IX, and X as well as proteins C and S (Horton and Bushwick, 1999). Green arrows depict activation, while red lines depict inhibition.
1.4.1 Extrinsic pathway

The main role of the extrinsic pathway is to generate a thrombin burst. Following damage to the blood vessel, FVII leaves the circulation and comes into contact with TF expressed on tissue-factor-bearing cells (stromal fibroblasts and leukocytes), forming an activated complex (TF-FVIIa) (Figure 1.3). TF-FVIIa activates factors IX and X. FVII is itself activated by thrombin, and factors Xla, XIIa and Xa. The activation of factor X (to form factor Xa) by TF-FVIIa is almost immediately inhibited by tissue factor pathway inhibitor (TFPI). Factor Xa and its co-factor factor Va form the prothrombinase complex, which activates prothrombin (factor II) to thrombin (factor IIa). Thrombin then activates other components of the coagulation cascade, including factors V and VIII, which activates factor XI. Factor XIa, and in turn, activates factor IX, which activates factor VIII releasing factor VIII from vWF.

1.4.2 Intrinsic pathway

Initiation of the intrinsic pathway occurs when prekallikrein (PK), high-molecular-weight kininogen (HMWK), factor XI and factor XII are exposed to the negatively charged phospholipids on the damaged vessel surface (contact phase) (Figure 1.3). The conversion of prekallikrein to kallikrein activates factor XII to factor XIIa. Factor XIIa then activates factor XI to factor Xla. Factor XIIa will also hydrolyse more prekallikrein to kallikrein, establishing a reciprocal activation cascade. In the presence of calcium (Ca\(^{2+}\)), factor Xla activates factor IX to factor IXa. The activation of factor Xa requires assemblage of the tenase complex (Ca\(^{2+}\) and factors VIIIa, IXa and X) on the surface of activated
platelets. The role of factor VIII in this process is to act as a receptor, in the form of factor VIIIa, for factors IXa and X. Factor VIIIa is termed a cofactor in the clotting cascade. The activation of factor VIII to factor VIIIa occurs in the presence of minute quantities of thrombin. As the concentration of thrombin increases, factor VIIIa is ultimately cleaved by thrombin and inactivated. This dual action of thrombin, upon factor VIII, acts to limit the extent of tenase complex formation and thus the extent of the coagulation cascade.

1.4.3 Common pathway

The common point in both the extrinsic and intrinsic pathways is the activation of factor X to factor Xa. Factor Xa activates prothrombin to thrombin, and thrombin in turns activates factor V to factor Va. Factor Va acts as a cofactor in the formation of the prothrombinase complex by binding to specific receptors on the surfaces of activated platelets and forms a complex with prothrombin and factor Xa (Figure 1.3).

Finally, thrombin converts fibrinogen to fibrin. Thrombin also activates factor XIII to factor XIIIa, a highly specific transglutaminase that introduces cross-links between the fibrin manomers, solidifying the fibrin clot. In addition to its role in activation of fibrin clot formation, thrombin plays an important regulatory role in coagulation. Thrombin combines with thrombomodulin present on endothelial cell surfaces forming a complex that converts protein C to activated protein C (APC). The cofactor protein S and protein APC degrade factors Va and VIIIa, thereby limiting their activity in the coagulation cascade.
1.4.4 Coagulation pathway regulators

Other regulators of the coagulation pathway are protein Z, TFPI and antithrombin III. Protein Z is a cofactor which forms a calcium ion-dependent complex with factor Xa at phospholipid surfaces enhancing the inhibition of factor Xa (Broze, 2001; Macfarlane, 1964). TFPI limits the action of tissue factor, thereby inhibits excessive TF-mediated activation of factors VII and X. Antithrombin III is a serine protease inhibitor that degrades the serine proteases: thrombin and factors IXa, Xa, XIa, and XIIa.

1.5 Warfarin

Warfarin (3-\(\alpha\)-acetylbenzyl-4-hydroxycoumadin) is a synthetic derivative of dicoumarol, a 4-hydroxycoumarin-derived mycotoxin anticoagulant originally discovered in spoiled sweet clover-based animal feed.

In the 1920s, cattle in North USA and Canada were afflicted by an outbreak of an unusual disease, characterised by fatal bleeding, either spontaneously or from minor injuries. This was later linked to their diet of mouldy sweet clover hay (\textit{Melilotus alba} and \textit{M. officinalis}) (Roderick, 1929; Schoefield, 1924) which was found to contain a haemorrhagic factor that reduced the activity of prothrombin (Roderick, 1931). However, it was not until 1940 that Karl Link and his student Harold Campbell in Wisconsin discovered that the anticoagulant in sweet clover was 3,3'-'-methylenebis (4-hydroxycoumarin) (Campbell and Link, 1941). Further work by Link led to the synthesis of warfarin in 1948, which was initially approved as a rodenticide in
the USA in 1952, and then for human use in 1954 (Clatanoff et al., 1954; Link, 1959; Shapiro, 1953). The name warfarin is derived from WARF (Wisconsin Alumni Research Foundation) and –arin from coumarin.

Warfarin has significant variability in pharmacological response among individuals, with doses varying between 1 to 20 mg per day (Johnson et al., 2011; Klein et al., 2009; Loebstein et al., 2001). In addition, warfarin has a narrow therapeutic window (Palareti et al., 1996), and variation in clinical response to this drug is consistently one of the leading causes of hospitalization for adverse drug events such as bleeding (Budnitz et al., 2011; Pirmohamed et al., 2004).

1.5.1 Clinical use of warfarin

Warfarin is one of the most widely prescribed oral anticoagulants worldwide (Kirley et al., 2012). It is indicated for the prophylaxis and treatment of thromboembolic complications associated with atrial fibrillation (AF), venous thrombosis, pulmonary embolism (PE), myocardial infarction (MI), and cardiac valve replacement. Table 1.1 outlines the common indications for warfarin therapy.

Over 25 million warfarin prescription have been made in the United States with world’s total prescriptions reaching 0.5-1.5% of the population (Johnson et al., 2011). However, because of the large difference in the dose requirements, the anticoagulation effect must be monitored regularly. Anticoagulation status is measured as the International Normalized Ratio (INR), which is a measure of the prothrombin time (PT) consisting of vitamin
K-dependent coagulation factors II, VII and X (van den Besselaar et al., 2004). INR is calculated by the following formula: \( \text{INR} = \frac{\text{patient's PT}}{\text{mean normal PT}} \times \text{ISI} \), where mean normal PT is the geometric mean of the PTs of the healthy adult population and ISI is the International Sensitivity Index of the thromboplastin used at the local laboratory to perform the PT measurement. A normal individual usually has an INR of 1. Warfarin is normally dosed to achieve an INR between 2.0 to 3.0 for most indications, except for patients with mechanical prosthetic heart valves and with systemic recurrent emboli (Table 1.1) (Hirsh et al., 1998; Kearon et al., 2012; Keeling et al., 2011). When a patient is started on warfarin, INR monitoring should be performed daily until the INR is within the therapeutic range for at least 2 consecutive days.

The safety and efficacy of warfarin therapy are dependent on maintaining the INR within the target range for the indication. Sub- and supra-therapeutic INR (INR<1.5 and INR>4.0, respectively) have been associated with an increased risk of thromboembolic and bleeding events, respectively (Johnson and Cavallari, 2013; Johnson et al., 2011). Vitamin K, an antagonist of warfarin, is generally used to reverse the effect of warfarin in the event of high INR or bleeding.
Table 1.1. Recommended INR therapeutic range for anticoagulant therapy*.

<table>
<thead>
<tr>
<th>Indication</th>
<th>INR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment of venous thrombosis</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>Treatment of pulmonary embolism</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>Prophylaxis of venous thrombosis (high-risk surgery)</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>Prevention of systemic embolism</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>Tissue heart valves</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>AMI (to prevent systemic embolism)</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>Valvular heart disease</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>Bileaflet mechanical valve in aortic position</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>Bioprosthetic valve in mitral position</td>
<td>2.0-3.0</td>
</tr>
<tr>
<td>Mechanical prosthetic valves (high risk)</td>
<td>2.5-3.5</td>
</tr>
<tr>
<td>Systemic recurrent emboli</td>
<td>2.5-3.5</td>
</tr>
</tbody>
</table>

*Recommended by the American College of Chest Physician (Ansell et al., 2008; Kearon et al., 2012).

#If anticoagulant therapy is elected to prevent recurrent myocardial infarction, an INR of 2.5-3.5 is recommended, consistent with recommendations of the Food and Drug Administration (FDA).

AMI, acute myocardial infarction; INR, international normalized ratio. Adapted from Hirsh et al., 1998.

Therapy with warfarin is usually started empirically on a fixed dose between 5 mg and 10 mg for the first 1 or 2 days for most individuals, with subsequent dosing based on the INR response. Older patients and those with liver disease, poor nutritional status, or heart failure may require lower initiation dosages (Kearon et al., 2012). The iterative process to define the appropriate dose can take weeks to months, and during this period patients are at increased risk of over- or under-anticoagulation (Budnitz et al., 2011; Pirmohamed, 2006; Pirmohamed et al., 2004).
Numerous factors are known to affect warfarin dose requirement in individuals (see below). Therefore, strategies to individualise warfarin therapy and reduce ADRs have been sought.

1.5.2 Pharmacokinetics of warfarin

Warfarin consists of a racemic mixture of two active optical isomers, the R- and S-enantiomers, in approximately equal proportions. Their pharmacokinetic and pharmacodynamic properties differ considerably, and they are cleared by different pathways. The S-warfarin accounts for 60-70% of the overall anticoagulant activity and is estimated to be 3 to 5 times more potent than the R-enantiomer (Breckenridge et al., 1974; O’Reilly, 1974).

Warfarin is highly water soluble, is rapidly absorbed from the gastrointestinal tract, and has high bioavailability (Breckenridge, 1978; O’Reilly, 1976). Warfarin can be detected in the plasma one hour after oral administration and the peak concentration occurs in two to eight hours after oral administration (Breckenridge, 1978; Kelly and O’Malley, 1979). Racemic warfarin has a half-life of 36-42 h (R-warfarin, 45 h; S-warfarin, 29 h) and the duration of the effect lasts up to five days (O’Reilly, 1987). After oral administration, 97-98% of racemic warfarin is bound to albumin, with 1% of free drug being pharmacologically active (Otagiri et al., 1987; Wilting et al., 1980).

Warfarin is metabolized in the liver and kidneys. Inactive metabolites are secreted in urine and stool. The human cytochrome P-450 (CYP) family plays a key role in warfarin metabolism, catalysing its hydroxylation at different
sites, and producing a series of mono-hydroxylated metabolites, through regio- and stereo-selective reactions (Brian et al., 1990; Hermans and Thijssen, 1989; Kaminsky et al., 1984, 1993; Nelson et al., 1996; Wang et al., 1983). R-warfarin is metabolised by CYP1A2, CYP2C19 and CYP3A4 to 6-, 7-, and 8-hydroxywarfarin that are excreted in the urine (Brian et al., 1990; Hermans and Thijssen, 1989; Kaminsky et al., 1984, 1993; Rettie et al., 1992; Wang et al., 1983; Zhang et al., 1995). S-warfarin is metabolized almost exclusively by CYP2C9 to 7-hydroxywarfarin and to a much lesser extent, 6-hydroxywarfarin (Kaminsky et al., 1984; Rettie et al., 1992; Wang et al., 1983). S-warfarin has about a 3-fold greater systemic clearance than R-warfarin (Breckenridge et al., 1974; O’Reilly, 1974; Wittkowsky, 2003).

1.5.3 Pharmacodynamics of warfarin

The procoagulant factors II, VII, IX, X, and anticoagulant proteins C, S and Z are vitamin K-dependant proteins which require γ-carboxylation of their glutamic acid residues to become fully functional (Ansell et al., 2004; Friedman et al., 1977; Malhotra et al., 1985; Stafford, 2005). The conversion of glutamic acid (Glu) residues on the N-terminal region of vitamin K-dependent proteins to γ-carboxyglutamic acid (Gla) residues induces a calcium-dependent conformational change in coagulation proteins which promotes their binding to cofactors on phospholipid surfaces (Borowski et al., 1986; Nelsestuen, 1976; Prendergast and Mann, 1977).

Carboxylation is catalysed by the vitamin K-dependent γ-carboxylation system, a multicomponent system of proteins that is embedded in the
membrane of the endoplasmic reticulum. It consists of the vitamin K-dependent γ-glutamyl carboxylase (GGCX), which requires the reduced hydroquinone form of vitamin K\textsubscript{1} (vitamin K\textsubscript{1}H\textsubscript{2}) as a cofactor, and the warfarin-sensitive enzyme vitamin K\textsubscript{1} 2,3-epoxide reductase (VKOR), which produces the cofactor. During the reaction, in the presence of molecular oxygen and carbon dioxide, the vitamin K\textsubscript{1}H\textsubscript{2} is oxidized to vitamin K\textsubscript{1} 2,3-epoxide, which is recycled to vitamin K by VKOR (Berkner, 2005; Wallin and Hutson, 2004). The vitamin K cycle is shown in Figure 1.4.

Warfarin exerts its anticoagulant effect by inhibiting VKOR, therefore interfering with the cyclic inter-conversion of vitamin K\textsubscript{1} and its 2,3-epoxide (Figure 1.3) (Choonara \textit{et al.}, 1988; Fasco \textit{et al.}, 1982; Nelsestuen \textit{et al.}, 1974; Stafford, 2005; Stenflo \textit{et al.}, 1974; Trivedi \textit{et al.}, 1988; Whitlon \textit{et al.}, 1978). This process leads to the depletion of vitamin K\textsubscript{1}H\textsubscript{2} and limits the γ-carboxylation of the vitamin K-dependent coagulant proteins (Wallin and Hutson, 2004). The reduction in the number of Gla residues from the normal 10–13 to less than 6 results in >95% decrease in coagulant activity (Malhotra, 1989, 1990; Malhotra \textit{et al.}, 1985).

When the liver concentration of vitamin K\textsubscript{1} is elevated, vitamin K\textsubscript{1} can be reduced to vitamin K\textsubscript{1}H\textsubscript{2} by an alternative enzyme, flavoprotein DT-diaphorase, which is a nicotinamide adenine dinucleotide phosphate (NAD(P)H) dehydrogenase (Hochstein, 1983; Wallin and Martin, 1987). Warfarin does not inhibit DT-diaphorase (Wallin and Martin, 1987). Therefore, warfarin intoxication can be counteracted by administering high doses of vitamin K\textsubscript{1} which drives cofactor vitamin K\textsubscript{1}H\textsubscript{2} production for GGCX, restoring normal functioning of the blood coagulation system.
Figure 1.4. Pharmacology of warfarin. Warfarin is administered as a racemic admixture of R- and S-enantiomers. The more potent S-enantiomer is metabolised principally by cytochrome P450 (CYP) 2C9. The pharmacological effect of warfarin is mediated by the inhibition of vitamin K epoxide reductase complex 1 (VKORC1). This results in decreased concentrations of activated clotting factors (II, VII, IX and X) producing therapeutic anticoagulation. GGCX, γ-glutamyl-carboxylase; CO₂, carbon dioxide; O₂, oxygen; HO₂, water.
1.6 Clinical and environmental factors affecting warfarin response

Clinical and environmental factors such as age, race, weight, height, medications, diet, illness, smoking and adherence all influence warfarin dosing.

Dose requirements decrease with age, owing to reduced clearance and/or increased responsiveness, by ~8-10% per decade of life (Gage et al., 2004; Gurwitz et al., 1992; Loebstein et al., 2001; Sconce et al., 2005). Mean warfarin daily dose requirements fell by 0.5 to 0.7 mg per decade between the ages of 20 to 90 years (Caldwell et al., 2007; Gage et al., 2004; Herman et al., 2006; Klein et al., 2009; Sconce et al., 2005; Wu et al., 2008).

Owing to differences in the prevalence of both known and unknown genetic variants, warfarin dose requirements vary by race. As compared with Caucasians, African-Americans require higher doses on average (Gage et al., 2004), whereas Asians require lower doses on average (Dang et al., 2005; Voora et al., 2005). However, analyses adjusting for known genetic factors, particularly VKORC1 genotype, have found that race is only a small contributor or is not a contributor to dose requirements (Gage et al., 2008; Klein et al., 2009).

Numerous concomitant medications can affect warfarin dose requirements, including amiodarone, antibiotics, salicylates, acetaminophen, thyroxine, inducers of cytochrome P450 (e.g. carbamazepine), HMG-CoA reductase inhibitors (e.g. simvastatin), and some herbal medicines (e.g. St John’s Wort and ginkgo) (D’Andrea et al., 2008; Gage et al., 2004; Moore et
al., 2000; Perini et al., 2008; Sanoski and Bauman, 2002; Wang et al., 2001). Table 1.2 lists some food and drugs that interact with warfarin.

Dietary factors such as alcohol consumption or vitamin K intake might influence warfarin dose requirements. Warfarin acts by inhibiting the recycling of vitamin K, and thus high intake of fat-soluble vitamin K (found in green vegetables) can theoretically diminish the action of warfarin. In addition, low or inconsistent vitamin K intake might contribute to variation in anticoagulation control (Loebstein et al., 2001; Sconce et al., 2005). Daily vitamin K supplementation for patients with previously unstable control was found to improve anticoagulation control (Sconce et al., 2007). However, there is conflicting evidence on the association between warfarin maintenance doses and vitamin K intake, where some studies did not find any association (Absher et al., 2002; Gage et al., 2004).

Various illnesses can affect warfarin dose requirements. Patients with liver disease, malnutrition, decompensated heart failure, hypermetabolic states (e.g. febrile illnesses, hyperthyroidism) and different therapeutic indications for warfarin (e.g. prosthetic heart valve, current venous thromboembolism) may lead to altered dose requirements (D’Andrea et al., 2008; Ansell et al., 2001; Gage and Eby, 2003; Gage et al., 2008).

Cigarette smoking has been demonstrated to induce CYP1A2, responsible for R-warfarin metabolism (Faber and Fuhr, 2004). With increased smoking, the activity of CYP1A2 increases, leading to increased metabolism of R-warfarin and a consequent need for higher warfarin dose requirement. Conversely, smoking cessation can cause an increase in warfarin activity. Therefore, changes in smoking behaviour could affect dose requirement.
Indeed careful monitoring of the INR in these patients is needed (Bourgeois et al., 2016; Bryk et al., 2015; Faber and Fuhr, 2004).

### Table 1.2. Some of major food and drug interactions with warfarin.

<table>
<thead>
<tr>
<th>Food</th>
<th>Herbal</th>
<th>Dietary supplement</th>
<th>Increased anticoagulation</th>
<th>Decreased anticoagulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chamomile</td>
<td>Dong quai</td>
<td>Garlic</td>
<td>Avocado</td>
<td>Broccoli</td>
</tr>
<tr>
<td>Herbal</td>
<td></td>
<td>Ginger</td>
<td>Co-Enzyme Q10</td>
<td>Green tea</td>
</tr>
<tr>
<td>Dietary</td>
<td></td>
<td>Gingko biloba</td>
<td></td>
<td>Spinach</td>
</tr>
<tr>
<td>supplement</td>
<td></td>
<td>Ginseng</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Guarana</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>St. John’s Wort</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drugs</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetaminophen</td>
<td>Ketoconazole</td>
<td>Antithyroid drugs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allopurinol</td>
<td>Metronidazole</td>
<td>Ascorbic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amiodarone</td>
<td>Miconazole</td>
<td>Azathioprine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Androgens</td>
<td>NSAIDs</td>
<td>Barbiturates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspirin</td>
<td>Omeprazole</td>
<td>Carbamazepine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azithromycin</td>
<td>Phenformin</td>
<td>Cholestyramine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Celecoxib</td>
<td>Phenylbutazone</td>
<td>Gluthimide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cimetidine</td>
<td>Phenytoin</td>
<td>Griseofulvin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ciprofloxacin</td>
<td>Probencid</td>
<td>Oral contraceptive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarithromycin</td>
<td>Quinidine</td>
<td>Rifampicin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clofibrate</td>
<td>Simvastatin</td>
<td>Sucralfate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>Sulfapyrazone</td>
<td>Vitamin K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dipyridamole</td>
<td>Tamoxifen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disulfiram</td>
<td>Thyroid hormone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erythromycin</td>
<td>Tobutamide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluconazole</td>
<td>Tramadol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluoroquinolones</td>
<td>Trimethoprim-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>sulfamethoxazole</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glucagon</td>
<td>Vancomycin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indomethacin</td>
<td>Verapamil</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Itraconazol</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*NSAIDs, nonsteroidal anti-inflammatory drugs.
1.7 Genetic contributions to warfarin response

Both candidate gene and GWAS data over the past decade have clearly shown that genetic polymorphisms contribute to the inter-patient variability in warfarin dose requirement (Aithal et al., 1999; Aquilante et al., 2006; Cha et al., 2010; Cooper et al., 2008; Rieder et al., 2005; Scordo et al., 2002; Takeuchi et al., 2009). More than 30 genes with a putative impact on warfarin anticoagulation have been identified (Jonas and McLeod, 2009). These genes have generally been found to have minimal impact on warfarin dose requirements. However, the variants of two genes, CYP2C9, which encodes the main warfarin metabolizing enzyme cytochrome P450 2C9, and VKORC1, which encodes for vitamin K epoxide reductase complex 1, have been shown to be the major genetic determinants influencing warfarin dose requirements.

The CYP4F2 gene, which encodes a vitamin K oxidase, has also been shown to influence warfarin pharmacodynamics and dose requirements, but to a lesser extent.

1.7.1 CYP2C9

The cytochrome P450 superfamily are the major enzymes involved in drug metabolism, accounting for approximately 75% of the total metabolism (Ingelman-Sundberg, 2004). The CYP2C9 gene is located on chromosome 10q24.2, spans approximatively 55 kb, contains 9 exons, and encodes a 60 kDa microsomal protein (Goldstein and de Morais, 1994).
CYP2C9 was the first gene documented to affect warfarin dose requirement (Furuya et al., 1995). Most of the variants in the CYP2C9 gene lead to reduced activity of the liver enzyme, resulting in increased sensitivity to warfarin (Niinuma et al., 2014). There are over 35 genetic variants in the CYP2C9 gene, of which the CYP2C9*2 (rs1799853) and *3 (rs1057910) are the most extensively studied (Table 1.3) (Crespi and Miller, 1997; Ieiri et al., 2000). CYP2C9*2 induces an Arg144Cys amino acid substitution in exon 3, whereas CYP2C9*3 encodes for Ile359Leu amino acid change in exon 7. S-warfarin clearance is reduced by approximately 40% with CYP2C9*2 and 75% with CYP2C9*3 (Scordo et al., 2002; Takahashi et al., 1998). Accordingly, warfarin dose requirements are approximately 20% lower with the CYP2C9*1/*2 genotype and 35% lower with the CYP2C9*1/*3 genotype compared with the CYP2C9*1/*1 genotype (Lindh et al., 2009). Doses of 1 mg/day or lower may be necessary in patients with the CYP2C9*3/*3 genotype to prevent over-anticoagulation and bleeding.

CYP2C9 allele frequencies differ by ancestry, with the CYP2C9*2 and *3 alleles occurring much more commonly among those of European ancestry versus Asian or African-American descents (Table 1.3) (Lam and Cheung, 2012; Perera et al., 2011). The CYP2C9*5 (rs28371686), *6 (rs9332131), *8 (rs7900194), and *11 (rs28371685) alleles predominate among those of African-American ancestry, with the *8 allele being the most common (Perera et al., 2011; Scott et al., 2009). Decreased S-warfarin clearance and lower warfarin dose requirements have been reported in patients bearing these alleles (Dickmann et al., 2001; Limdi et al., 2008a; Liu et al., 2011b; Perera et al., 2011; Tai et al., 2005).
Overall, CYP2C9 genotype explains approximately 10 to 15% of the interpatient variability in warfarin dose requirements (Anderson et al., 2007; Aquilante et al., 2006; Carlquist et al., 2006; Cavallari et al., 2010; Klein et al., 2009; Sconce et al., 2005; Wadelius et al., 2007; Wu et al., 2008). CYP2C9 variant alleles are also associated with an increased risk for over-anticoagulation and bleeding during warfarin therapy (Aithal et al., 1999; Limdi et al., 2008b). The risk of bleeding attributable to CYP2C9 polymorphisms is highest during the initial months of warfarin therapy. However, there is evidence that this persists during chronic therapy, suggesting a need for close monitoring for signs and symptoms of bleeding throughout warfarin therapy for carriers of the CYP2C9 variant alleles (Limdi et al., 2008b).

1.7.2 VKORC1

SNPs in the VKORC1 gene have been linked to reduced efficacy in vitamin K recycling as a result of lower VKOR activity (Rost et al., 2004a). The VKORC1 gene, approximately 4 kb long, is located on chromosome 16p11.2 (Li et al., 2004; Rost et al., 2004a).

Genetic variants in the VKORC1 coding region have been shown to be associated with warfarin resistance, where very high doses (>20 mg/day) were required to obtain therapeutic anticoagulation (Rost et al., 2004a). VKORC1 variants associated with warfarin resistance are rare in most populations, with the exception of the Ashkenazi Jewish population, in whom the p.Asp36Try variant occurs at a prevalence of approximately 8% (Scott et al., 2008).
In the general population, SNPs in the *VKORC1* promoter (-1639G>A, rs9923231) and intron 1 (1173C>T, rs9934438) have been associated with warfarin sensitivity, and they are responsible for approximately 25% of the warfarin dose variability (Rieder *et al*., 2005; Wang *et al*., 2008). However, -1639G>A and 1173C>T occur in near complete linkage disequilibrium across populations, and thus, either may be considered for warfarin dose prediction (Limdi *et al*., 2010). The -1639AA, AG, and GG genotypes (or 1173TT, CT, and CC genotypes) are associated with high, intermediate, and low sensitivity to warfarin, respectively. Thus, when compared with the -1639AG genotype, higher warfarin doses are needed for carriers of the GG genotype to effectively inhibit vitamin K reduction and subsequent clotting factor activation, whereas carriers of the AA genotype require lower doses.

Another common *VKORC1* SNP, the 3730G>A (rs7294), located in the 3’ untranslated region of the gene, has been associated with increased warfarin dose requirements (Geisen *et al*., 2005; Rieder *et al*., 2007). Using linear regression models, this polymorphism has been shown to account for about 10% of warfarin dose requirements in the Italian population (Borgiani *et al*., 2009; Cini *et al*., 2012).

As shown in Table 1.3, the -1639A allele frequency varies by ancestry, and largely explains the differences in warfarin dose requirements between Caucasian, African-American, and Asian populations. Specifically, the African populations have a higher frequency of the GG genotype and generally require higher warfarin doses than those of European descent. Asians have a high frequency of the AA genotype and usually require very low doses of warfarin. The *VKORC1* -1639G>A variant explains approximately 20 to 35% of warfarin
dose variability in Caucasians and Asians, but only 5 to 7% in African-Americans (Aquilante et al., 2006; Cavallari et al., 2010; Limdi et al., 2010). The lesser variability explained in African-Americans is primarily attributable to the lower -1639A allele frequency in this group (Limdi et al., 2010). Nonetheless, persons of African descent derive a similar benefit from genotype-guided warfarin dosing as others.

1.7.3 CYP4F2

CYP4F2 metabolizes vitamin K to hydroxyl-vitamin K, resulting in less vitamin K available for clotting factor activation. The CYP4F2 gene is located on chromosome 19p13.12. The activity of CYP4F2 is reduced in individuals with the CYP4F2 p.Val433Met SNP (rs2108622, c.1297G>A), resulting in reduced vitamin K metabolism and greater vitamin K availability (McDonald et al., 2009). Studies in both Caucasians and Asians have shown higher warfarin dose requirements with the CYP4F2 AA genotype (Caldwell et al., 2008; Cha et al., 2010; Gong et al., 2011a; Sagrieya et al., 2010; Wei et al., 2012). Due to the lower allele frequency in African-Americans, the association of warfarin dose with the polymorphism is not observed in this population (Bress et al., 2012; Cha et al., 2010). In GWAS conducted in European and Asian populations, CYP4F2 emerges as a predictor for warfarin dose requirements after controlling for CYP2C9 and VKORC1, explaining an additional 1 to 3% of the overall variability (Cha et al., 2010; Takeuchi et al., 2009). Although this SNP has been demonstrated to be significantly associated with warfarin dose
requirements, the effect size is very small and its clinical use remains to be further studied.

1.7.4 Other genetic factors

In addition to CYP2C9, VKORC1 and CYP4F2, more than 30 genes in the warfarin interactive pathways have been studied as potential contributors to the variation of warfarin doses (see Table 1.4) (Cavallari et al., 2010; Cha et al., 2007; Chen et al., 2005; Kimura et al., 2007; Lal et al., 2008; Lee et al., 2009; Loebstein et al., 2005; Otagiri et al., 1987; Rieder et al., 2007; Shikata et al., 2004; Wadelius and Pirmohamed, 2007; Wadelius et al., 2004, 2005, 2007; Wajih et al., 2004). Figure 1.5 illustrates an overview of warfarin interactive pathways whereas Table 1.4 details protein functions.

SNPs in the gene encoding γ-glutamyl carboxylase (GGCX) have been reported to have an influence (albeit minimal) on dose requirements in some studies (Chen et al., 2005; Crawford et al., 2007; Kimura et al., 2007; Shikata et al., 2004; Wadelius et al., 2005). Cavallari et al. (2012) reported an association of the GGCX SNP rs10654848 with higher warfarin dose requirements in African-Americans. The GGCX SNP explained 2% of warfarin dose variability and occurs 10 times more frequently in African-Americans than in Caucasians (MAF 2.6% versus 0.27%) (Cavallari et al., 2012).

The SNP rs339097 in calumenin (a chaperon protein capable of inhibiting GGCX) has been demonstrated to confer 11-15% higher warfarin dose requirements in African-Americans (Voora et al., 2010). The minor allele
frequency of rs339097 is approximately 1% in Europeans as opposed to 25% in African-Americans (Cavallari et al., 2011).

The apolipoprotein E (APOE) gene has been evaluated in Caucasians, Africans and Asians (Kimmel et al., 2008; Kohnke et al., 2005; Lal et al., 2008). Plasma vitamin K levels vary by APOE genotype (Saupe et al., 1993). There are three common variants of the APOE gene (designated ε2, ε3 and ε4) and their prevalence varies by ethnicity. The ε4 variant is more common in African-Americans (21%) (Kimmel et al., 2008) than in Caucasians (13%) (Kohnke et al., 2005) and Asians (11%) (Lal et al., 2008). APOE genotypes were found to affect warfarin dose requirements in some studies, but not in others, suggesting that the possible effect of APOE may be population specific (Caldwell et al., 2007; Huang et al., 2011; Kimmel et al., 2008; Kohnke et al., 2005; Lal et al., 2008; de Oliveira Almeida et al., 2014).

A SNP in the CYP2C18 gene (rs12777823), was also shown to affect warfarin dose requirements in a population of African-American ancestry (Perera et al., 2013). Carriers of the minor A allele had required lower warfarin doses, suggesting reduced clearance of S-warfarin.
Table 1.3. Reported prevalence of \textit{CYP2C9}, \textit{VKORC1}, and \textit{CYP4F2} gene polymorphisms by ancestry (Bress \textit{et al.}, 2012; Cavallari \textit{et al.}, 2010; Cha \textit{et al.}, 2010; Chan \textit{et al.}, 2012; Gage \textit{et al.}, 2008; Limdi \textit{et al.}, 2008a, 2008b, 2010; Perera \textit{et al.}, 2013).

<table>
<thead>
<tr>
<th>SNP</th>
<th>Allele</th>
<th>Prevalence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2C9*2 (rs1799853)</td>
<td>*2</td>
<td>24</td>
</tr>
<tr>
<td>CYP2C9*3 (rs1057910)</td>
<td>*3</td>
<td>12</td>
</tr>
<tr>
<td>CYP2C9*5 (rs28371686)</td>
<td>*5</td>
<td>&lt;1</td>
</tr>
<tr>
<td>CYP2C9*6 (rs9332131)</td>
<td>*6</td>
<td>&lt;1</td>
</tr>
<tr>
<td>CYP2C9*8 (rs7900194)</td>
<td>*8</td>
<td>NR</td>
</tr>
<tr>
<td>CYP2C9*11 (rs28371685)</td>
<td>*11</td>
<td>&lt;1</td>
</tr>
<tr>
<td>VKORC1 -1639G&gt;A (rs9923231)</td>
<td>G</td>
<td>61</td>
</tr>
<tr>
<td>CYP4F2 1297G&gt;A (rs2108622)</td>
<td>A</td>
<td>40</td>
</tr>
</tbody>
</table>

*Percentage of individuals who have the variant. Adapted from Johnson and Cavallari, 2013.
Figure 1.5. An overview of warfarin interactive pathways. The figure illustrates genes thought to be involved in the action and biotransformation of warfarin and vitamin K. Adapted from Wadelius et al., 2007.
Table 1.4. Other genes involved in warfarin mechanism of action.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Gene</th>
<th>Function of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biotransformation of warfarin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Transport</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha-1-acid glycoprotein 1, Orosomucoid 1</td>
<td>ORM1</td>
<td>A plasma glycoprotein that functions as a carrier of warfarin in the blood (Nakagawa et al., 2003; Otagiri et al., 1987)</td>
</tr>
<tr>
<td>Alpha-1-acid glycoprotein 2, Orosomucoid 2</td>
<td>ORM2</td>
<td>A plasma glycoprotein that functions as a carrier of warfarin in the blood (Nakagawa et al., 2003; Otagiri et al., 1987)</td>
</tr>
<tr>
<td>P-glycoprotein, Multidrug resistance protein 1</td>
<td>ABCB1 (MDR1)</td>
<td>A cellular efflux pump for xenobiotics (Kroetz et al., 2003). Warfarin is a weak inhibitor and maybe a substrate.</td>
</tr>
<tr>
<td><strong>Metabolism</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytocrome P450 1A1</td>
<td>CYP1A1</td>
<td>Metabolism of R-warfarin (Grossman et al., 1993; Kaminsky and Zhang, 1997; Zhang et al., 1995)</td>
</tr>
<tr>
<td>Cytocrome P450 1A2</td>
<td>CYP1A2</td>
<td>Metabolism of R-warfarin (Kaminsky and Zhang, 1997; Zhang et al., 1995)</td>
</tr>
<tr>
<td>Cytocrome P450 1A6</td>
<td>CYP1A6</td>
<td>Metabolism of S-warfarin? (Freeman et al., 2000)</td>
</tr>
<tr>
<td>Cytocrome P450 2C8</td>
<td>CYP2C8</td>
<td>Minor pathway for R- and S-warfarin (Kaminsky and Zhang, 1997; Rettie et al., 1992)</td>
</tr>
<tr>
<td>Cytocrome P450 2C18</td>
<td>CYP2C18</td>
<td>Minor pathway for R- and S-warfarin (Kaminsky and Zhang, 1997; Kaminsky et al., 1993)</td>
</tr>
<tr>
<td>Cytocrome P450 2C19</td>
<td>CYP2C19</td>
<td>Minor pathway for R- and S-warfarin (Kaminsky and Zhang, 1997; Kaminsky et al., 1993)</td>
</tr>
<tr>
<td>Cytocrome P450 3A4</td>
<td>CYP3A4</td>
<td>Metabolism of R-warfarin (Kaminsky and Zhang, 1997)</td>
</tr>
<tr>
<td>Cytocrome P450 3A5</td>
<td>CYP3A5</td>
<td>Metabolism of R-warfarin? (Huang et al., 2004)</td>
</tr>
<tr>
<td><strong>Cytochrome P450 inducibility</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pregnane X receptor (PXP)</td>
<td>NR112</td>
<td>Mediates induction of CYP2C9, CYP3A4, other CYP enzymes and ABCB1 (Chen et al., 2004; Geick et al., 2001; Lehmann et al., 1998)</td>
</tr>
<tr>
<td>Constitutive androstane receptor (CAR)</td>
<td>NR113</td>
<td>Transcriptional regulation of genes including CYP2C9 and CYP3A4 (Assenat et al., 2004)</td>
</tr>
<tr>
<td><strong>Biotransformation of vitamin K</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Transport</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein E</td>
<td>APOE</td>
<td>Apolipoprotein E serves as a ligand for receptors that mediate the uptake of vitamin K (Berkner and Runge, 2004; Kohlmeier et al., 1996; Lamon-Fava et al., 1998; Sause et al., 1993)</td>
</tr>
<tr>
<td><strong>Vitamin K cycle</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epoxide hydrolase 1, microsomal</td>
<td>EPHX1</td>
<td>A hepatic epoxide hydrolase in the endoplasmic reticulum that may be complexed with VKOR (Cain et al., 1997; Loebstein et al., 2005; Morisseau and Hammock, 2005)</td>
</tr>
<tr>
<td>NAD(P)H dehydrogenase, quinone 1</td>
<td>NQO1</td>
<td>A detoxifying enzyme that has the potential to reduce the quinine form of vitamin K (Berkner and Runge, 2004; Ross and Siegel, 2004; Wallin and Hutson, 1982)</td>
</tr>
<tr>
<td>Calumenin</td>
<td>CALU</td>
<td>Binds to the vitamin K epoxide reductase complex and inhibits the effect of warfarin (Wajih et al., 2004; Wallin et al., 2001)</td>
</tr>
<tr>
<td>ß-glutamyl carboxylase</td>
<td>GGCX</td>
<td>Carboxylates vitamin-K-dependent coagulation factors and proteins in the vitamin K cycle (Berkner, 2000; Rost et al., 2004b; Wu et al., 1997)</td>
</tr>
</tbody>
</table>
### Table 1.4. Continued

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Gene</th>
<th>Function of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vitamin K-dependent proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coagulation factor II, prothrombin</td>
<td>F2</td>
<td>Converts fibrinogen to fibrin, activates F5, FVIII, FXI, FXIII, protein C (Berkner, 2000; Dahlbäck, 2005)</td>
</tr>
<tr>
<td>Coagulation factor VII</td>
<td>F7</td>
<td>Is converted to FVIIa and then converts FIX to FXa and FX to FXa (Berkner, 2000; Dahlbäck, 2005)</td>
</tr>
<tr>
<td>Coagulation factor IX</td>
<td>F9</td>
<td>Makes a complex with FVIIIa and then converts FX to FXa (Berkner, 2000; Dahlbäck, 2005)</td>
</tr>
<tr>
<td>Coagulation factor X</td>
<td>F10</td>
<td>Converts FII to FIIa in the presence of FVa (Berkner, 2000; Dahlbäck, 2005)</td>
</tr>
<tr>
<td>Protein C</td>
<td>PROC</td>
<td>Activated protein C counteracts coagulation together with protein S by inactivating FVa and FVIIIa (Berkner, 2000; Dahlbäck, 2005)</td>
</tr>
<tr>
<td>Protein S</td>
<td>PROS1</td>
<td>Cofactor to protein C that degrades FVa and FVIIIa (Berkner, 2000; Dahlbäck, 2005)</td>
</tr>
<tr>
<td>Protein Z</td>
<td>PROZ</td>
<td>Is together with protein Z-dependent protease inhibitor, a cofactor for the inactivation of FXa (Berkner, 2000; Broze, 2001)</td>
</tr>
<tr>
<td>Growth-arrest-specific protein 6</td>
<td>GAS6</td>
<td>Participates in many processes, i.e. potentiation of agonist-induced platelet aggregation (Berkner and Runge, 2004)</td>
</tr>
<tr>
<td><strong>Other coagulation proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-thrombin III</td>
<td>SERPINC1</td>
<td>Inhibits FIIa, FIXa, FXa, FXia and FXIIa. Anti-thrombin deficiency increases risk of thrombosis (Dahlbäck, 2005)</td>
</tr>
<tr>
<td>Coagulation factor V</td>
<td>F5</td>
<td>A cofactor that activates FII together with FXa. An F5 mutation leads to risk of thrombosis (Dahlbäck, 2005)</td>
</tr>
</tbody>
</table>

* Adapted from Wadelius and Pirmohamed, 2007.
1.8 Warfarin dosing algorithms

Due to the difficulties in determining the required stable warfarin dose for a given patient, many different regression models for dose prediction have been proposed worldwide to improve anticoagulation management, reduce complications, and enhance efficacy. These models vary in terms of the predictive factors they include, with some including clinical variables, such as age, body weight, gender, concomitant medications and indication for warfarin therapy (Ovesen et al., 1989; Theofanous and Barile, 1973; Williams and Karl, 1979). Other models include details of initial INR measurements and loading warfarin doses (Le Gal et al., 2010; Solomon et al., 2004). The usage of computer programs that incorporate clinical variables has been shown to be helpful in maintaining a more stable INR and reducing complication rates in both the induction and maintenance phases during warfarin therapy (Ageno et al., 2000; Manotti et al., 2001).

More recently, warfarin dosing algorithms have also included genetic factors which have been shown to be associated with dosing requirements. Some of the first dosing algorithms incorporating CYP2C9 genotype were published in 2004 (Gage et al., 2004; Hillman et al., 2004; Kamali et al., 2004). The algorithm by Gage and colleagues (2004) was the most extensive and included, in addition to CYP2C9 genotype, age, body surface area (BSA), gender, race, target INR, amiodarone use and simvastatin use, explaining 39% of the variability in warfarin dose requirement.

Since then, more than 30 warfarin dosing algorithms for various ethnic groups have been published based on both CYP2C9 and VKORC1 genotypes (Anderson et al., 2007; Avery et al., 2011; Carlquist et al., 2006; Cho et al., 2008;...
2011; Cini et al., 2012; Gage et al., 2008; Herman et al., 2006; Horne et al., 2012; Huang et al., 2009c; Klein et al., 2009; Lenzini et al., 2010; Pathare et al., 2012; Pavani et al., 2012; Perini et al., 2008; Ramos et al., 2012; Sasaki et al., 2009; Sconce et al., 2005; Suriapranata et al., 2011; Takahashi et al., 2006; Tham et al., 2006; Wadelius et al., 2009; Wu et al., 2008; You et al., 2011). More recently, several algorithms that included CYP4F2 polymorphisms have been developed (Borgiani et al., 2009; Choi et al., 2011; Gong et al., 2011b; Harada et al., 2010; Wei et al., 2012; Wells et al., 2010; Xu et al., 2012; Zambon et al., 2011).

Given the enormous evidence supporting the association of genetic factors with warfarin dose requirement and the regulatory mandate for safety, in August 2007, the FDA approved the addition of pharmacogenetic data to the warfarin labelling. The label states that lower doses “should be considered for patients with certain genetic variations in CYP2C9 and VKORC1 enzymes”.

Overall, demographic factors (such as age, gender, height and weight) and the use of concomitant medications account for approximately 10-15% of warfarin dose variability, whereas genetic factors (mostly CYP2C9 and VKORC1) account for an additional 35-50% of the warfarin dose variance (Figure 1.6).
Figure 1.6. Factors contributing to the inter-individual variation in warfarin dose requirements. Numerous clinical and genetic factors have been associated with variability in warfarin dose requirement. Adapted from Jonas and McLeod, 2009.

However, the problem with most of these algorithms is that they were developed using a relatively small sample size usually from a single ethnic group. As such, the International Warfarin Pharmacogenetics Consortium (IWPC) developed a highly reliable warfarin dosing algorithm in a large and diverse population (4,043 patients) from nine countries. The IWPC algorithm, incorporating information on CYP2C9 and VKORC1 genotypes, age, height, weight, amiodarone use, race and number of CYP enzyme inducers, predicted 47% of warfarin dose variation among patients.

Based on these findings, in January 2010 the FDA made a second revision to the warfarin labelling to include a dosing table that included CYP2C9 and VKORC1 genotypes (Table 1.5). The table may be used as a quick guide for clinicians to dose warfarin when genotype is available, realizing that clinical factors still need to be taken into account. Pharmacogenetic information is also included in the warfarin labelling approved by the Pharmaceuticals and Medical Devices Agency in Japan and by regulatory body in Taiwan.
Table 1.5. Recommended warfarin starting dose (mg/day) according to VKORC1 and CYP2C9 genotypes (FDA-approved warfarin labelling).

<table>
<thead>
<tr>
<th>VKORC1 -1639</th>
<th>CYP2C9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*1/*1</td>
</tr>
<tr>
<td>GG</td>
<td>5-7</td>
</tr>
<tr>
<td>AG</td>
<td>5-7</td>
</tr>
<tr>
<td>AA</td>
<td>3-4</td>
</tr>
</tbody>
</table>

1.9 miRNAs as potential biomarkers of therapy response

Identified in 1993 (Lee et al., 1993; Wightman et al., 1993), microRNAs (miRNAs) are a class of evolutionarily conserved, small non-coding RNAs of 17-24 nucleotides in length that modulate cellular messenger RNA (mRNA) and protein levels by interacting with specific mRNAs, usually at the 3’ untranslated region (UTR), through partial sequence complementation (Ambros, 2004; Zhao and Liu, 2009). miRNAs represent approximatively 1-2% of known genes in eukaryotes (Bartel, 2004; John et al., 2004), and participate in a broad range of processes including cell cycling, apoptosis, cell differentiation, tumour development, invasion, metastasis and angiogenesis (Friedman et al., 2009). Thus far, over 2,500 human miRNAs have been identified (www.mirbase.org) (Griffiths-Jones et al., 2006; Kozomara and Griffiths-Jones, 2011, 2014).

Recently, a significant number of miRNAs have also been found outside of the cells, and the levels of some of these extracellular miRNAs in
circulation have been linked to different pathophysiological conditions, such as cancer, myocardial infarction, and drug-induced liver injury (Adachi et al., 2010; Mitchell et al., 2008; Wang et al., 2010, 2009; Weber et al., 2010). These findings raise the possibility of using the circulating levels of specific miRNAs as biomarkers for different pathological conditions (Creemers et al., 2012; Etheridge et al., 2011; Kosaka et al., 2010; Laterza et al., 2009; Li et al., 2010a; Reid et al., 2011; Schwarzenbach et al., 2011; Wang et al., 2009).

From a biological point of view, miRNAs may be better predictive and prognostic markers than DNA or mRNA. A single miRNA may regulate hundreds of target mRNAs, frequently grouped in specific biological pathways. Consequently, a miRNA signature may provide prognostic information that is several orders of magnitude greater than mRNAs. Besides, miRNAs are very stable, and are thus more suitable for analysis in plasma, urine and stool (Cortez and Calin, 2009). In the case of large-scale screening, the possibility of studying markers from easily accessible and inexpensive surrogate tissues like plasma and stool could be ideal. In addition, plasma miRNAs were found to be remarkably stable even under conditions as harsh as boiling, low or high pH, long-time storage at room temperature, and multiple freeze-thaw cycles (Chen et al., 2008; Lawrie et al., 2008; Mitchell et al., 2008). Indeed, several studies have shown that miRNAs can be (i) packaged in microparticles such as exosomes, microvesicles, and apoptotic bodies (Fleissner et al., 2012; Valadi et al., 2007; Zernecke et al., 2009), or (ii) associated with the RNA-binding protein Argonaute2 (Arroyo et al., 2011; Turchinovich et al., 2011) or (iii) lipoprotein complexes such as high-density lipoprotein (HDL) (Vickers et al., 2011) to prevent their degradation.
Recently, miRNAs have been proposed to play a role in warfarin dose requirement (Ciccacci et al., 2015; Pérez-Andreu et al., 2012), suggesting that there remains substantial room for improvement in the understanding of factors predicting warfarin dose. Further studies to evaluate the contribution of miRNAs in warfarin dosing might be the key to improve warfarin dose prediction from pharmacogenetics algorithms.

1.10 Thesis aims

Although warfarin is an effective anticoagulant, determining the dose required to achieve a stable therapeutic INR is difficult due to the large inter-individual variability and its narrow therapeutic index. A recent meta-analysis of randomised controlled trials provided reliable evidence for the decreased risks of major bleeding and thromboembolic events in warfarin-treated patients guided by pharmacogenetic dosing algorithms (Li et al., 2015). Nevertheless, each algorithm may be biased when applied to a population other than that from which it is derived, mainly because of the genetic differences between the studied populations.

Warfarin dosing algorithms developed in the Italian population (Borgiani et al., 2009; Cini et al., 2012; Zambon et al., 2011) have included two SNPs, VKORC1 3730G>A (rs7294), and CYP4F2 1297G>A (rs2108622), which were found to be significantly associated with warfarin dose requirements in the Italian population. However, none of these studies included patients from Southern Italy.
Furthermore, all the algorithms published to date, at best, explain about 70% of warfarin dose variance (Verhoef et al., 2014). Other novel factors yet to be identified could explain some of the missing variations.

Given that previous studies suggested a possible role for miR-133a in warfarin therapy response (Ciccacci et al., 2015; Pérez-Andreu et al., 2012; Shomron, 2010), circulating miRNAs could possibly explain part of the large warfarin inter-patient variability.

Further studies need to be addressed to identify other genetic and non-genetic factors that better predict warfarin dosage requirements.

The aims of this PhD study were therefore (i) to develop an algorithm to predict the individual maintenance dose of warfarin in patients from Southern Italy on the basis of clinical and genetic data, and to compare the prediction accuracy of our pharmacogenetic dosing algorithm with other algorithms from the literature; (ii) to evaluate the potential effects of circulating miRNAs on warfarin therapy response by conducting a pilot study comparing the baseline serum levels of miRNA in patients who have achieved warfarin stable dose to those who did not achieve warfarin stability; (iii) to perform a genome-wide association study in patients requiring low or high doses of warfarin to identify additional variants contributing to warfarin dose sensitivity or resistance.
CHAPTER 2

Pharmacogenetic-based warfarin dosing algorithm in Southern Italians
CONTENTS

2.1 INTRODUCTION ........................................................................................................... 47

2.2 MATERIALS AND METHODS ................................................................................. 49
  2.2.1 Subjects .................................................................................................................. 49
  2.2.2 Samples and genotyping ...................................................................................... 50
    2.2.2.1 TaqMan Genotyping and allelic discrimination ............................................ 50
    2.2.2.2 Denaturing high performance liquid chromatography .............................. 53
    2.2.2.3 Genotyping quality control by Sanger sequencing ..................................... 56
  2.2.3 Statistical analysis ............................................................................................... 57
    2.2.3.1 Univariate analysis ....................................................................................... 57
    2.2.3.2 Multiple regression analysis ....................................................................... 58
    2.2.3.3 Correction for multiple testing .................................................................... 58
    2.2.3.4 Evaluation of our warfarin dosing algorithm ............................................. 59
    2.2.3.5 Dose prediction accuracy comparison between our
          pharmacogenetic dosing algorithm and published
          pharmacogenetic algorithms ............................................................................. 59

2.3 RESULTS .................................................................................................................... 63
  2.3.1 Patient characteristics ......................................................................................... 63
  2.3.2 Univariate analysis of association between clinical variables and
      weekly warfarin stable dose ..................................................................................... 65
  2.3.3 Univariate analysis of association between SNPs and weekly
      warfarin stable dose ............................................................................................... 71
  2.3.4 Multiple regression models ................................................................................ 78
  2.3.5 Evaluation of the pharmacogenetic algorithm and comparison
      with published algorithms ..................................................................................... 82

2.4 DISCUSSION .............................................................................................................. 86
2.1 INTRODUCTION

Numerous warfarin dosing algorithms integrating genetics with clinical and environmental factors have been developed. However, the performance of the different algorithms is highly variable, explaining between 31 to 79% of warfarin dose requirements (Kim et al., 2009b; Lubitz et al., 2010). This large variability is due to differences in study design, the set of predictive variables considered and the ethnicity of populations evaluated. Interethnic variability is related to the diversity of the genetic background of patients and, in particular, to the different allelic frequencies in CYP2C9, VKORC1 and CYP4F2 (D’Andrea et al., 2005; Borgiani et al., 2007, 2009; Caldwell et al., 2008; Limdi and Veenstra, 2008; Schelleman et al., 2008; Takahashi et al., 2006).

In a genome-wide analysis of 126K autosomal SNPs within the Italian population, Di Gaetano et al. (2012) demonstrated that the genetic structure of the Italian population was strongly influenced by geographical location. Northern Italians are genetically closer to the French population, whereas Southern Italians share genetic similarities with the Mediterranean populations, including those from the Middle East. In addition, some degree of North African ancestry is observed across South West Europe (Di Gaetano et al., 2012). Interestingly, in a previous study on postmenopausal women from Southern Italy, we found that the minor allele frequency of a SNP in the vitamin D receptor (VDR) gene, rs2228570 (C>T), was different from that described in women from Central Italy and other Caucasian populations, but is similar to that in the African-American and Maltese populations (Conti et al., 2015).

To date, several warfarin dosing algorithms have been developed in the Italian population (Borgiani et al., 2009; Cini et al., 2012; Zambon et al.,
2011), but none of them included patients from Southern Italy. The aims of this experimental chapter therefore were to: (i) develop an algorithm to predict the individual maintenance dose of warfarin in patients from Southern Italy on the basis of clinical and genetic data, evaluating in particular the contribution of CYP2C9 (*2 and *3), CYP4F2 (1297G>A), VKORC1 (-1639 G>A, 1173 C>T and 3730 G>A) SNPs to warfarin dose requirement; and (ii) validate and compare the prediction accuracy of our pharmacogenetic dosing algorithm with those from the literature, including the pharmacogenetic algorithm developed by the International Warfarin Pharmacogenetics Consortium (IWPC) (Klein et al., 2009) and pharmacogenetic algorithms developed in the Italian (Borgiani et al., 2009; Cini et al., 2012; Zambon et al., 2011) and American-Caucasian (Anderson et al., 2007) populations.
CHAPTER 2

2.2 MATERIALS AND METHODS

2.2.1 Subjects

Two hundred and sixty-six warfarin-treated patients of Caucasian descent from Southern Italy were retrospectively recruited from three hospitals in Southern Italy: (i) University of Naples Federico II Hospital, Department of Internal Medicine, (ii) Salvatore Maugeri Foundation, IRCCS, Scientific Institute of Terme, Benevento, and (iii) Santobono Pausilipon Hospital, Naples (Mazzaccara et al., 2013). Inclusion criteria for the study were ≥ 18 years of age and stable anticoagulation. Stable anticoagulation was defined as an unchanged daily dose of warfarin for at least 3 months where INR measurements were within the individual’s target range, according to Higashi et al. (Higashi et al., 2002). Exclusion criteria were acenocoumarol therapy and known non-compliance with oral anticoagulant therapy. All subjects provided written informed consent to participate in the study. The study was approved by the research ethics committees from the University of Naples Federico II, the Foundation Salvatore Maugeri IRCCS Institute of Campoli Telesse, Benevento, and the Second University of Naples, and was performed according to the second Helsinki Declaration.

Data on clinical factors that have previously been shown to be associated with warfarin therapy were recorded on a structured interview form. These data included demographic characteristics, primary indication for warfarin treatment, stable therapeutic dose of warfarin, target INR, smoking status, comorbidities and the use of concomitant medications (grouped according to those that increase and those that decrease INR). Information on
hypertension, systolic blood pressure above 130 mmHg and diastolic blood pressure above 85 mmHg, and body mass index (body weight [kg] divided by squared height [m$^2$]) were also recorded. Liver dysfunction (aspartate aminotransferase >35 U/L women, >40 U/L men; alanine aminotransferase >35 U/L women, >40 U/L men), and dyslipidaemia (serum total cholesterol and/or triglycerides levels above 190 mg/dL or 150 mg/dL, respectively), were also measured.

2.2.2 Samples and genotyping

Genomic DNA was extracted from EDTA-stabilized blood using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer’s instructions.

2.2.2.1 TaqMan Genotyping and allelic discrimination

Genotyping of $CYP2C9^{*2}$ (rs1799853), $CYP2C9^{*3}$ (rs1057910), $CYP4F2$ 1297G>A (rs2108622), and $VKORC1$ -1639 G>A (rs9923231) were performed using TaqMan® SNP Genotyping assays (Applied Biosystems, Foster City, CA), according to the manufacturer’s instructions. This procedure is based on the 5’–3’ exonuclease activity of Taq DNA polymerase, using allele-specific TaqMan® minor groove binding (MGB) probes labelled with two fluorescent reporter dyes (VIC® and FAM™) that target a SNP site (Figure 2.1). Each fluorescent dye is specific to one allele. Fluorescence associated with the probes accumulates as the number of PCR cycles increases.
Approximately 10-20 ng genomic DNA was amplified in a 25 µL reaction mixture in a 96-well plate containing 1X universal TaqMan® genotyping master mix and 1X assay mix containing a premix of the respective primers and fluorescent-labelled MGB probes. As part of quality control, negative controls containing water instead of DNA, positive controls with DNA of known genotype, and 10% duplicates were included in every run.

The reactions were performed on an ABI Prism 7900-HT real-Time PCR system (Applied Biosystems Foster City, CA). End-point fluorescence and allelic discrimination were determined using the SDS version 2.1 software (Applied Biosystems Foster City, CA).
Figure 2.1. TaqMan® Probe-based chemistry for allelic discrimination. TaqMan® assays contain two specific primers targeting the region flanking the SNP site and two TaqMan fluorescent probes with a Minor Groove Binder (MGB). Each probe is labelled with a different fluorophore (VIC® or FAM™). These reporters are attached covalently to the 5’ end of the two probes. Near the 3’ end, there is a non-fluorescent quencher (Q) that prevents liberation of the reporter fluorescence if the probe is not degraded. The MGB serves to stabilize the double stranded structure formed between the target and the probe. During the PCR reactions, probes that hybridize specifically to DNA fragments are destroyed and the fluorescence of corresponding fluorophore is liberated. A substantial increase in VIC fluorescence only or FAM fluorescence only indicates homozygosity for one of the two alleles, while an increase in both fluorescent signals indicates heterozygosity.
2.2.2.2 Denaturing high performance liquid chromatography

Two variants, \textit{VKORC1} 1173C>T (rs9934438) and \textit{VKORC1} 3730G>A (rs7294) were detected by denaturing high performance liquid chromatography (DHPLC) on Wave 2.0 Transgenomic instruments (Transgenomic Inc., Omaha, Nebraska, USA). DHPLC is a chromatographic technique for the separation and analysis of DNA fragments with different length and/or base composition. The key to DHPLC is the solid phase which has differential affinity for single and double-stranded DNA. In DHPLC, DNA fragments are denatured by heating and then allowed to re-anneal. The melting temperature of the re-annealed DNA fragments determines the length of time they are retained in the column. Figure 2.2 summarises the principles of allelic discrimination by DHPLC.

PCR was performed in a reaction volume of 30 µL, containing approximately 50 ng of genomic DNA in 0.2 mL tubes. Each reaction mixture contained 67 mM Tris-\(\text{HCl}\) (pH 8.8), 16.0 mM (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4}, 5 mM MgCl\textsubscript{2}, 0.2 mM dNTPs, 0.4 µM of oligonucleotide primers and 1U of EuroTaq DNA Polymerase (Euroclone, Milan, Italy). Forward and reverse primers for the two \textit{VKORC1} SNPs are listed in Table 2.1. Thermal cycling was performed in a Cycler TM Thermal Cycler (Bio-Rad, Richmond, CA) at the following amplification conditions: initial denaturation for 5 min at 95°C, followed by 35 PCR cycles of 30 seconds denaturation at 94°C, 90 seconds annealing at 60°C, and 30 seconds extension at 72°C; and a final extension step for 5 min at 72°C. The size and specificity of the PCR products was then examined in
2% ethidium bromide-stained tris-borate EDTA (TBE) agarose gel under UV light.

To allow heteroduplex formation, PCR products were subjected to 3 min 95°C denaturation followed by a gradual re-annealing from 95°C to 40°C in 30 min in a Cycler TM Thermal Cycler (Bio-Rad, Richmond, CA). The optimal melting temperature was based on recommendations from the Navigator® software, version 2.0 (Transgenomic Inc., Omaha, Nebraska, USA) and the DHPLC Melt program (Jones et al., 1999) (http://insertion.stanford.edu/melt.html).

Finally, re-annealed PCR products (5 µL) were injected onto a DNA Sep HT Column (Transgenomic Inc., Omaha, NE, USA), eluted on a linear acetonitrile gradient (0.1 M triethylammonium acetate (TEAA)/0.1 M TEAA with 25% acetonitrile). Samples were run on a constant flow rate of 0.9 mL/min under partially denaturing conditions at 60°C, and hetero- and homoduplex were determined by UV absorbance at 260 nm.

DHPLC analysis was performed using the Navigator® software, version 2.0 (Transgenomic Inc., Omaha, Nebraska, USA). Interpretation of DHPLC data was based on the comparison between sample and reference chromatograms (controls of known genotype).

Samples which had chromatograms with unexpected or noisy peaks were then sequenced (Sanger sequencing, see below, section 2.2.2.3). In addition, as part of quality control to avoid misclassification of genotypes, 10% of randomly selected samples were also genotyped by Sanger sequencing.
Figure 2.2. Theoretical basis for DHPLC. Using PCR, two fragments are generated; target DNA containing the SNP polymorphic site and an allele-specific DNA sequence, referred to as the normal DNA fragment. This normal fragment is identical to the target DNA except at the SNP polymorphic site, which is unknown in the target DNA. The fragments are denatured and then allowed to gradually re-anneal. The re-annealed products are added to the DHPLC column. If the SNP allele in the target DNA matches the normal DNA fragment, only identical homoduplexes will form during the re-annealing step. If the target DNA contains a different SNP allele to the normal DNA fragment, heteroduplexes of the target DNA and normal DNA containing a mismatched polymorphic site will form. The mismatched heteroduplexes will have a different melting temperature to the homoduplexes and will not be retained in the column as long. This generates a chromatographic pattern that is distinctive from the pattern that would be generated if the target DNA fragment and normal DNA fragments were identical. Examples of chromatogram are shown on the right side of the figure.
Table 2.1. Primers for *VKORC1* 1173C>T and 3730G>A genotyping by DHPLC.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Primer sequence (5’→3’)</th>
<th>Primer direction</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>VKORC1</em> 1173C&gt;T</td>
<td>TGACATGGAATCCTGACGTG</td>
<td>Forward</td>
</tr>
<tr>
<td></td>
<td>TAGTGTGTAGAGATGCAACC</td>
<td>Reverse</td>
</tr>
<tr>
<td><em>VKORC1</em> 3730G&gt;A</td>
<td>GCTTTGCTTTGGCATGTGAG</td>
<td>Forward</td>
</tr>
<tr>
<td></td>
<td>ACACATGGTTCAGACTTGGC</td>
<td>Reverse</td>
</tr>
</tbody>
</table>

2.2.2.3 Genotyping quality control by Sanger sequencing

PCR products generated for DHPLC heteroduplexing (see section 2.2.2.2, paragraph 2) were used for Sanger sequencing. 5 µl of PCR product was purified with the enzyme ExoSAP-IT® (USB Corporation, Cleveland, Ohio, USA) for 15 min at 37°C in 1.5 x reaction buffer to degrade primers and dephosphorylate dNTPs that were not consumed in the reaction. The reaction was stopped by a 15 min incubation at 80°C.

A sequencing reaction was performed with 1 µl of purified PCR product using the BigDye Terminator version 1.1 Cycle Sequencing Kit and ABI 3100 Genetic Analyser (both, Applied Biosystems Foster City, CA), according to the standard protocol. The internal forward sequence for each SNP was used as the sequencing primer.

Sequencing reactions and sequence data analyses were performed by Dr. Cristina Mazzaccara at CEINGE Advanced Biotechnology s.c.a.r.l., Naples.
2.2.3 Statistical analysis

Each SNP was tested for deviation from Hardy–Weinberg Equilibrium (HWE) using the exact test in the PLINK software, version 1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al., 2007). A P<0.001 was assumed to indicate deviation from HWE.

Analysis of linkage disequilibrium (Gabriel et al., 2002) was performed using the Haploview software, version 4.2 (Barrett et al., 2005).

The Kolmogorov-Smirnov test was performed to evaluate the distribution of the outcome stable warfarin dose. As the distribution of warfarin stable dose was skewed, square-root transformation was undertaken to ensure normal distribution for the purpose of downstream association analyses.

Data were expressed as average ± standard deviation (SD) (continuous variables) or in percentage (categorical variables).

Statistical analysis was performed with the Statistical Package for the Social Sciences (SPSS 22.0, SPSS Inc., Chicago, IL, USA).

2.2.3.1 Univariate analysis

To test the differences in clinical and demographic features among patients, the Student t test and analysis of variance (ANOVA) were undertaken for continuous variables. For categorical variables, the Pearson's chi-square test was used.

To evaluate the effects of clinical variables and SNPs on the continuous outcome of stable warfarin dose, two univariate tests of association
were conducted: one-way analysis of variance (ANOVA) which makes no assumption on the mode of inheritance, and univariate linear regression which assumes an additive mode of inheritance. A $P<0.05$ was considered statistically significant.

### 2.2.3.2 Multiple regression analysis

Clinical and genetic factors which showed significant univariate associations ($P_c<0.1$) were included in the multiple regression analyses. To avoid collinearity, correlation between each pair of clinical factors, as well as LD between each pair of SNPs, were assessed, with only one from each highly correlated pair represented in the model. To determine the relative effects of the clinical and genetic factors on the variability of warfarin stable dose, stepwise linear regression models were fitted and compared using the likelihood ratio test (LRT). A pharmacogenetics algorithm was then derived and the predicted warfarin dose for each patient was calculated using the regression equation.

### 2.2.3.3 Correction for multiple testing

To account for multiple testing, the false discovery rate (FDR) (Benjamini et al., 2001), was calculated in the genetics package of R, version 3.1.1 (http://cran.rproject.org/web/packages/genetics/index.html) in addition to the raw $P$-value. In calculating the FDR, $P$-values from all association tests undertaken were taken into account. FDR-corrected $P$-values are denoted as $P_c$-values. A $P_c<0.1$ was regarded as statistically significant.
2.2.3.4 **Evaluation of our warfarin dosing algorithm**

In order to compare the predicted dose to the observed weekly warfarin dose, dose units were standardised where all square-root transformations were converted to mg/week. To evaluate the predictive accuracy of our algorithm, the mean absolute error (MAE) which is the average of differences between the predicted and observed doses, was calculated.

The clinical accuracy of our algorithm was evaluated by calculating the percentage of patients whose predicted dose was within 20% of the actual stable therapeutic dose.

2.2.3.5 **Dose prediction accuracy comparison between our pharmacogenetic dosing algorithm and published pharmacogenetic algorithms**

Using the following five published dosing algorithms, the predicted weekly warfarin dose for each of our patients was calculated:

- **Zambon et al., 2011**

  \[
  \text{Square-root of weekly warfarin dose (mg) = } 7.398 - 0.027 \times \text{age (years)} + 1.063 \times \text{body surface area (m}^2) - 1.045 \times (\text{VKORC1 -1639 GA}) - 2.121 \times (\text{VKORC1 -1639 AA}) - 0.790 \times (\text{CYP2C9 *1/*2}) - 1.171 \times (\text{CYP2C9 *1/*3}) - 1.813 \times (\text{CYP2C9 *2/*2, *2/*3, *3/*3}) - 0.467 \times (\text{CYP4F2 GA}) - 0.715 \times (\text{CYP4F2 GG}).
  \]

  Age and body surface area were continuous variables; VKORC1 -1639 GG, CYP2C9*1/*1 and CYP4F2 AA were the reference genotypes in this pharmacogenetic algorithm. The model predicts
the square root of weekly dose. Hence to obtain dose in mg/week, the result was squared.

- Cini et al., 2012

Square-root of daily warfarin dose (mg) = 0.833 – 0.255 (male gender) – 0.007 x age (years) + 0.011 x height (cm) + 0.005 x weight (kg) + 0.385 (smoking) + 0.211 (vegetable intake) + 0.119 (indication for venous thromboembolism) + 0.328 (diabetes) – 0.458 (variant alleles of CYP2C9) – 0.571 (VKORC1 -1639G>A) + 0.025 (VKORC1 3730G>A).

Variables were defined as the following: gender (female=0, male=1); age, height, weight were continuous variables; smoking (no=0, yes=1), vegetable intake, indication for venous thromboembolism (no=0, yes=1), diabetes (no=0, yes=1); variant alleles of CYP2C9 (none=0, one-variant allele (i.e. *1/*2 or *1/*3) =1, two-variant alleles (i.e. *2/*2, or *2/*3 or *3/*3) =2); VKORC1 -1639G>A and VKORC1 3730G>A (GG=0, GA=1, AA=2). We did not collect information on vegetable intake in our study; hence a value of 0 was therefore assigned to all patients. The model predicts the square root of daily dose. Thus to obtain dose in mg/week, the result was squared and multiplied by seven.

- Borgiani et al., 2009

Weekly warfarin dose (mg) = 35.0 – 0.34 x age (years) + 0.25 x weight (kg) – 0.25 (variant alleles of CYP2C9*2) – 0.23 (variant
alleles of CYP2C9*3) + 0.26 (variant alleles of CYP4F2 1297A) – 0.32 (VKORC1 1173C>T) + 0.9 (VKORC1 3730G>A).

Variables were defined as the following: age, weight (continuous variables); variant alleles of CYP2C9*2, CYP2C9*3 and CYP4F2 1297A (none=0, one-variant allele=1, two-variant alleles=2); VKORC1 1170C>T and VKORC1 3730G>A (GG=0, GA=1, AA=2).

- Couma-Gen algorithm (Anderson et al., 2007)

\[
\text{Weekly warfarin dose (mg)} = 35.0 - 0.321 \times \text{age (years)} + 0.135 \times \text{weight (kg)} - 0.101 \times (\text{male gender}) - 0.232 \times \text{CYP2C9 *1/*2} - 0.206 \times \text{CYP2C9 *1/*3} - 0.155 \times \text{CYP2C9 *3/*3} - 0.331 \times \text{VKORC1 1173 CT} - 0.447 \times \text{VKORC1 1173 TT}.
\]

Age and weight were continuous variables; VKORC1 1173 TT and CYP2C9*1/*1 were the reference genotype in this pharmacogenetic algorithm.

- IWPC algorithm (Klein et al., 2009)

\[
\text{Square-root of weekly warfarin dose (mg)} = 5.6044 - 0.2614 \times \text{age in decades} + 0.0087 \times \text{height (cm)} + 0.0128 \times \text{weight (kg)} - 0.8677 \times \text{VKORC1 -1639 GA} - 1.6974 \times \text{VKORC1 -1639 AA} - 0.4854 \times \text{VKORC1 -1639 genotype unknown} - 0.5211 \times \text{CYP2C9 *1/*2} - 0.9357 \times \text{CYP2C9 *1/*3} - 1.0616 \times \text{CYP2C9 *2/*2} - 1.9206 \times \text{CYP2C9 *2/*3} - 2.3312 \times \text{CYP2C9 *3/*3} - 0.2188 \times \text{CYP2C9 genotype unknown} + 1.1816 \times \text{enzyme inducer status} - 0.5503 \times \text{amiodarone status (no=0, yes=1)}.
\]
Age in decades (1 for 10-19, 2 for 20-29, etc); height and weight were continuous variables; VKORC1 -1639 GG and CYP2C9*1/*1 were the reference genotype in this pharmacogenetic algorithm; enzyme inducer status=1 if patients taking phenytoin, phenobarbital, or rifampicin, otherwise 0; amiodarone status (no=0, yes=1). Since all our patients were of Caucasian descent from Southern Italy, the contribution of the race was not taken into account. The model predicts the square root of weekly dose; hence the dose (mg/week) was obtained by squaring the result.

To assess the prediction accuracy of the above 5 dosing algorithms, the MAE was calculated and compared as mg/week across all models. The proportion of variance explained by each model was evaluated by calculating the coefficient of determination ($R^2$). In addition, the clinical accuracy of these algorithms was explored by calculating the percentage of patients whose predicted dose fell within 20% of the observed dose; the predicted warfarin dose was also correlated with the observed dose using Pearson correlation.
2.3 RESULTS

2.3.1 Patient characteristics

The studied population consisted of 266 warfarin-treated patients of Caucasian descent from Southern Italy (55.3% male, mean age 67.13±11.30 years). All patients were retrospectively recruited.

Complete dosage information was available for all the patients. Data on adverse events such as thromboembolic or bleeding events were not available.

The clinical and demographic features of patients are shown in Table 2.2. Cardiac valve replacement and atrial fibrillation were the most frequent indications for warfarin therapy (42.5% and 36.8%, respectively).
Table 2.2. Clinical and demographic features of our warfarin-treated patients (n=266).

<table>
<thead>
<tr>
<th>Overall population</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>67.13 ± 11.30</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>55.3%</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>26.94 ± 4.27</td>
</tr>
<tr>
<td>BSA (m²)</td>
<td>1.86 ± 0.18</td>
</tr>
<tr>
<td>Warfarin stable dose (mg/wk)</td>
<td>28.57 ± 13.31</td>
</tr>
<tr>
<td>Indication for warfarin therapy</td>
<td></td>
</tr>
<tr>
<td>Cardiac valve replacement</td>
<td>42.5%</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>36.8%</td>
</tr>
<tr>
<td>Dilatative cardiomyopathy</td>
<td>8.3%</td>
</tr>
<tr>
<td>Deep venous thrombosis</td>
<td>7.1%</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>2.6%</td>
</tr>
<tr>
<td>Other</td>
<td>2.6%</td>
</tr>
<tr>
<td>INR therapeutic target (range)</td>
<td></td>
</tr>
<tr>
<td>2.5 (2.0-3.0)</td>
<td>64.3%</td>
</tr>
<tr>
<td>3.0 (2.5-3.5)</td>
<td>35.7%</td>
</tr>
<tr>
<td>Smoking</td>
<td>8.6%</td>
</tr>
<tr>
<td>Dyslipidaemia</td>
<td>65.0%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>62.8%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>15.4%</td>
</tr>
<tr>
<td>Liver disease</td>
<td>15.4%</td>
</tr>
<tr>
<td>Kidney failure</td>
<td></td>
</tr>
<tr>
<td>Mild renal impairment</td>
<td>4.9%</td>
</tr>
<tr>
<td>Severe renal impairment</td>
<td>3.4%</td>
</tr>
<tr>
<td>Concomitant drug assumption</td>
<td></td>
</tr>
<tr>
<td>Enzyme inducer status</td>
<td>1.9%</td>
</tr>
<tr>
<td>Phenytoin</td>
<td>0.4%</td>
</tr>
<tr>
<td>Phenobarbital</td>
<td>1.5%</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>0.4%</td>
</tr>
<tr>
<td>Enzyme inhibitor status</td>
<td>19.2%</td>
</tr>
<tr>
<td>Amiodarone</td>
<td>10.9%</td>
</tr>
<tr>
<td>Omeprazole</td>
<td>6.0%</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>4.1%</td>
</tr>
</tbody>
</table>

Data are expressed as average ± standard deviation (SD) (continuous variables) or in percentage (categorical variables).

BMI, body mass index; BSA, body surface area. Other indications for warfarin therapy included: transient ischaemic attack (TIA), stroke. Enzyme inducer status: patients taking phenytoin, rifampicin, or phenobarbital; Enzyme inhibitor status: patients taking amiodarone, omeprazole, or simvastatin.
2.3.2 Univariate analysis of association between clinical variables and weekly warfarin stable dose

Table 2.3 summarises the results of the univariate analyses between all clinical variables and our outcome of warfarin stable dose. Of the 25 variables tested, significant associations were found with age ($P_c=1.457 \times 10^{-3}$), BSA ($P_c=0.067$), cardiac valve replacement as indication for warfarin therapy ($P_c=0.090$), and amiodarone status ($P_c=0.026$).

This finding is consistent with several studies (Klein et al., 2009; Wadelius et al., 2009; Zambon et al., 2011), although some authors demonstrated an association of male gender with higher warfarin requirement (Absher et al., 2002; Anderson et al., 2007; Cini et al., 2012).

The correlations of weekly warfarin stable dose with age and BSA are shown in Figures 2.3A and 2.3B, respectively. Decreased weekly warfarin dose was observed as age increased ($R=-0.279$) while a direct correlation was seen between warfarin dose and BSA ($R=0.143$).

Patients who underwent warfarin therapy after cardiac valve replacement required a higher weekly dose ($P_c=0.090$, Figure 2.4A). Indeed, the target therapeutic INR range for these patients was much higher (2.5-3.5 versus 2.0-3.0; Pearson Chi-Square, $P=3.305 \times 10^{-23}$). Interestingly, when all patients were taken into account, no significant difference in dose requirements was observed between the two groups of patients with different target therapeutic INR ranges ($P_c=0.734$; Figure 2.4C).

To avoid collinearity, correlation between variables which showed significant univariate associations was assessed. A significant correlation between cardiac valve replacement and age was found ($P=0.123 \times 10^{-3}$; Figure
2.4B). Given that age showed a stronger association with warfarin stable dose, it was subsequently included in the regression model.

Patients in concomitant therapy with amiodarone, an inhibitor of cytochrome P450 enzymes, required a significantly lower dose of warfarin to achieve stable anticoagulation ($P_c=0.045$; Figure 2.4D). Indeed, several studies have reported that patients on amiodarone require nearly 30% lower doses of warfarin to achieve stable anticoagulation (Gage et al., 2008; Klein et al., 2009; Sanoski and Bauman, 2002; Santos et al., 2014)
Table 2.3. Univariate analyses for the continuous outcome of stable warfarin dose (square-root transformed).

<table>
<thead>
<tr>
<th></th>
<th>P-value</th>
<th>Pc-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>0.004x10^{-3}</td>
<td>1.457x10^{-5}</td>
</tr>
<tr>
<td>Gender</td>
<td>0.258</td>
<td>0.417</td>
</tr>
<tr>
<td>BMI (kg/m^2)</td>
<td>0.660</td>
<td>0.727</td>
</tr>
<tr>
<td>BSA (m^2)</td>
<td>0.023</td>
<td>0.067</td>
</tr>
<tr>
<td>Indication for warfarin therapy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac valve replacement</td>
<td>0.037</td>
<td>0.090</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>0.209</td>
<td>0.358</td>
</tr>
<tr>
<td>Dilatative cardiomyopathy</td>
<td>0.994</td>
<td>0.827</td>
</tr>
<tr>
<td>Deep venous thrombosis</td>
<td>0.062</td>
<td>0.139</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>0.645</td>
<td>0.727</td>
</tr>
<tr>
<td>Other</td>
<td>0.912</td>
<td>0.805</td>
</tr>
<tr>
<td>INR therapeutic target (range)</td>
<td>0.756</td>
<td>0.734</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.905</td>
<td>0.805</td>
</tr>
<tr>
<td>Dyslipidaemia</td>
<td>0.593</td>
<td>0.727</td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.781</td>
<td>0.734</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.965</td>
<td>0.827</td>
</tr>
<tr>
<td>Liver disease</td>
<td>0.702</td>
<td>0.727</td>
</tr>
<tr>
<td>Kidney failure</td>
<td>0.272</td>
<td>0.417</td>
</tr>
<tr>
<td>Mild renal impairment</td>
<td>0.724</td>
<td>0.727</td>
</tr>
<tr>
<td>Severe renal impairment</td>
<td>0.118</td>
<td>0.245</td>
</tr>
</tbody>
</table>

Concomitant drug assumption

Enzyme inducer status

- Phenytoin: 0.141 (Pc=0.274)
- Phenobarbital: 0.209 (Pc=0.358)
- Rifampicin: 0.293 (Pc=0.427)

Enzyme inhibitor status

- Amiodarone: 0.008 (Pc=0.026)
- Omeprazole: 0.461 (Pc=0.640)
- Simvastatin: 0.576 (Pc=0.727)

FDR-corrected P-values are denoted as Pc-values. Any variable having a significant univariate association (Pc<0.1) was designated as a candidate for multiple regression analysis.

Other indications for warfarin therapy included transient ischaemic attack (TIA) and stroke.
Figure 2.3. Correlation between warfarin stable dose and (A) age and (B) BSA (Body Surface Area). The linear correlation coefficient (R), the P-value (Pearson correlation), and the $P_c$-value after FDR correction are shown in the upper right corner.
Figure 2.4. Relationship between (A) indication for warfarin treatment and weekly stable dose, (B) indication for warfarin treatment and age, (C) warfarin stable dose and INR therapeutic target, and (D) warfarin stable dose and amiodarone status. Box and whisker plots showing the distribution data. Boxes represent the 25th-75th percentiles (interquartile range), whiskers represent 5th-95th percentiles, solid lines represent the median values in each group and dots represent outliers. FDR-corrected $P_c$-values are shown in the upper right corner.
2.3.3 Univariate analysis of association between SNPs and weekly warfarin stable dose

All six genotyped SNPs CYP2C9*2 (rs1799853), CYP2C9*3 (rs1057910), CYP4F2 1297G>A (rs2108622), VKORC1 -1639G>A (rs9923231), VKORC1 1173C>T (rs9934438), and VKORC1 3730G>A (rs7294) were in Hardy–Weinberg equilibrium (CYP2C9*2, \(P=0.799\); CYP2C9*3, \(P=0.116\); CYP4F2 1297G>A, \(P=0.250\); VKORC1 -1639G>A, \(P=0.806\); VKORC1 1173C>T, \(P=0.497\); VKORC1 3730G>A, \(P=0.756\)).

The allele and genotype frequencies for CYP2C9*2, CYP2C9*3, CYP4F2 1297G>A and VKORC1 (3730G>A and 1173C>T) were similar to those reported in other Caucasian populations (see Table 2.4) (D’Andrea et al., 2005; Borgiani et al., 2007, 2009; Chang et al., 2015; Cini et al., 2012; Di Fusco et al., 2013; Limdi et al., 2008a; Sánchez-Diz et al., 2009; Scott et al., 2010; Zambon et al., 2011), except for VKORC1 -1639 where the A allele showed higher prevalence in our population.

The clinical and demographic features of warfarin-treated patients stratified by genotypes of the six SNPs investigated are shown in Appendices 1.1 and 1.2.

The distribution of weekly warfarin stable doses per genotype group for each SNP are presented in Figure 2.5. Among the six genotyped SNPs, only four showed significant association with weekly warfarin stable dose in the univariate analysis: CYP2C9*2 (\(P_c=0.090\)), CYP2C9*3 (\(P_c=1.789\times10^{-7}\)), VKORC1 -1639G>A (\(P_c=2.491\times10^{-16}\)), and VKORC1 3730G>A (\(P_c=1.457\times10^{-5}\)).
Patients bearing the CYP2C9 *1/*2, and *2/*2 genotypes required significantly lower warfarin doses than patients with the wild-type genotype (25.27±10.90 mg/week, and 21.70±4.89 mg/week respectively versus 29.87±13.98 mg/week) (Figure 2.5A).

Warfarin dose requirement was also significantly lower in CYP2C9*3 homozygotes (10.78±9.78 mg/week) and heterozygotes (20.42±9.27 mg/week) than in patients with the wild-type genotype (30.27±13.24 mg/week) (Figure 2.5B).

The variant A allele of VKORC1 -1639G>A SNP (Figure 2.5C) was associated with lower dose requirement (18.82±7.99 mg/week, homozygous AA; 29.12±11.79 mg/week, heterozygous GA; 36.90±14.20 mg/week, wild-type GG).

For VKORC1 3730G>A, higher warfarin dosages were required by heterozygous patients (32.36±13.49 mg/week) and patients with the homozygous variant genotype AA (33.39±16.08 mg/week) when compared to patients carrying the wild-type genotype GG (24.38±11.12 mg/week) (Figure 2.5E).

No significant difference in warfarin dosages was observed with CYP4F2 1297G>A (Figure 2.5C) and VKORC1 1173C>T (Figure 2.5D) polymorphisms.
Table 2.4a. Allele and genotype frequencies of *CYP2C9*, and *CYP4F2* 1297G>A polymorphisms in the studied population and in other Caucasian populations.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Genotype frequencies</th>
<th>Allelic frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Our population</td>
<td>Other studies*</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td><strong>CYP2C9</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*1/*1</td>
<td>160</td>
<td>60.1</td>
</tr>
<tr>
<td>*1/*2</td>
<td>58</td>
<td>21.8</td>
</tr>
<tr>
<td>*1/*3</td>
<td>33</td>
<td>12.4</td>
</tr>
<tr>
<td>*2/*2</td>
<td>6</td>
<td>2.3</td>
</tr>
<tr>
<td>*2/*3</td>
<td>5</td>
<td>1.9</td>
</tr>
<tr>
<td>*3/*3</td>
<td>4</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>CYP4F2 1297G&gt;A</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G/G</td>
<td>121</td>
<td>45.5</td>
</tr>
<tr>
<td>G/A</td>
<td>123</td>
<td>46.2</td>
</tr>
<tr>
<td>A/A</td>
<td>22</td>
<td>8.3</td>
</tr>
</tbody>
</table>
Table 2.3b. Allele and genotype frequencies of VKORC1 (-1639G>A, 1173C>T, 3730G>A) polymorphisms in the studied population and in other Caucasian populations.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Genotype frequencies</th>
<th>Allelic frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Our population</td>
<td>Other studies*</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>VKORC1 -1639G&gt;A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G/G</td>
<td>69</td>
<td>26.0</td>
</tr>
<tr>
<td>G/A</td>
<td>131</td>
<td>49.2</td>
</tr>
<tr>
<td>A/A</td>
<td>66</td>
<td>24.8</td>
</tr>
<tr>
<td>VKORC1 1173C&gt;T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>114</td>
<td>42.9</td>
</tr>
<tr>
<td>C/T</td>
<td>116</td>
<td>43.6</td>
</tr>
<tr>
<td>T/T</td>
<td>36</td>
<td>13.5</td>
</tr>
<tr>
<td>VKORC1 3730G&gt;A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G/G</td>
<td>128</td>
<td>48.1</td>
</tr>
<tr>
<td>G/A</td>
<td>112</td>
<td>42.1</td>
</tr>
<tr>
<td>A/A</td>
<td>26</td>
<td>9.8</td>
</tr>
</tbody>
</table>

* Other studies: D’Andrea et al., 2005; Borgiani et al., 2007, 2009; Chang et al., 2015; Cini et al., 2012; Di Fusco et al., 2013; Limdi et al., 2008a; Sánchez-Diz et al., 2009; Scott et al., 2010; Zambon et al., 2011.
Figure 2.5. Influence of CYP2C9, CYP4F2, and VKORC1 genotypes on the warfarin stable dose. Box and whisker plots showing the distribution of warfarin weekly doses based on (A) CYP2C9*2, (B) CYP2C9*3, (C) CYP2F2 1297G>A, (D) VKORC1 -1639G>A, (E) 1173C>T and (F) 3730G>A genotypes. Boxes represent the 25th-75th percentiles (interquartile range), whiskers represent 5th-95th percentiles, solid lines represent the median value of warfarin weekly dose in each group and dots represent outliers. FDR-corrected \( P_c \)-values are shown in the upper right corner.
LD between each pair of SNPs in the *VKORC1* gene was assessed (Figure 2.6), and, consistently with other studies (Carlquist *et al.*, 2010; Cini *et al.*, 2012; Wadelius *et al.*, 2005), the analysis showed a lack of LD between *VKORC1*-1639G>A and *VKORC1* 3730G>A (*D'* = 0.785, *r*^2^ = 0.254).

![LD plot of the 3 VKORC1 SNPs, VKORC1 -1639G>A (rs9923231), VKORC1 1173C>T (rs9934438) and VKORC1 3730G>A (rs7294). Gabriel *et al.*’s definition (2002) was used to define the LD block in Haploview 4.2. VKORC1 -1639G>A (rs9923231) is not in strong linkage disequilibrium with VKORC1 3730G>A (rs7294) (*D'* = 0.785, *r*^2^ = 0.254).]
2.3.4 Multiple regression models

Using stepwise linear regression, we first developed a model including clinical factors which incorporated age, BSA and amiodarone status. The combination of these three clinical factors explained 16.5% of warfarin dose variability.

We then developed a second model incorporating genetic factors only. Of the four genetic factors (CYP2C9*2, CYP2C9*3, VKORC1 -1639G>A, VKORC1 3730G>A) which showed significant univariate association, only three (CYP2C9*2, CYP2C9*3, VKORC1 -1639G>A) were fitted in the model. These three SNPs attributed to 40.5% of warfarin dose variability.

To develop a pharmacogenetic algorithm, we incorporated both clinical (age, BSA, amiodarone status) and genetic factors (CYP2C9*2, CYP2C9*3, VKORC1 -1639G>A) in the model. The combination of these clinical and genetic factors accounted for 53.6% of warfarin dose variability. When compared to the clinical algorithm, the addition of genetic factors in the pharmacogenetic algorithm explained a further 37.1% of warfarin dose variability (LRT \( P \)-value = \( 1.340 \times 10^{-26} \)).

Table 2.5 summarises the relative effects of clinical and genetic factors on warfarin dose variability as described above.

From our pharmacogenetic algorithm, the following regression equation was obtained and was used to calculate the predicted warfarin dose:

\[
\text{Square-root of weekly warfarin dose (mg)} = 6.156 - 0.029 \text{ x age (years)} + 1.114 \text{ x body surface area (m}^2\text{)} - 0.632 \text{ (VKORC1 -1639 GA)} - 1.627 \text{ (VKORC1 -1639 AA)} - 0.350 \text{ (CYP2C9 *1/*2)} - 0.850
\]
(CYP2C9 *1/*3) – 0.662 (CYP2C9 *2/*2 or *2/*3) – 3.376 (CYP2C9 *3/*3) – 0.594 x amiodarone status (no=0, yes=1).

VKORC1 -1639 GG and CYP2C9 *1/*1 were the reference genotypes in our pharmacogenetic algorithm. Due to the small number of CYP2C9 *2/*2 and *2/*3 carriers in our cohort of patients, the CYP2C9 *2/*2 and *2/*3 genotypes were pooled. Several studies (Anderson et al., 2007; Cini et al., 2012; Roper et al., 2010) have suggested that the number of variant alleles within one gene, regardless of the type of variants, could correlate with therapeutic dose in a linear manner, with statistically significant dose differences across groups.

Our model predicts the square root of weekly dose. As such, the dose in mg/week is obtained by squaring the result. Table 2.6 outlines the parameter estimates of the variables included in our pharmacogenetic algorithm.
Table 2.5. Contribution of variables to warfarin stable dose requirements.

<table>
<thead>
<tr>
<th>Covariates</th>
<th>$P_{c}$-value</th>
<th>$P_{c}$-value LRT</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical only (age, BSA, amiodarone status)</td>
<td>2.088x10^{-7}</td>
<td>16.5</td>
<td></td>
</tr>
<tr>
<td>Genetics only ($CYP2C9^*2$, $CYP2C9^*3$, $VKORC1$ -1639G&gt;A)</td>
<td>6.132x10^{-28}</td>
<td>40.5</td>
<td></td>
</tr>
<tr>
<td>Clinical + genetics</td>
<td>1.842x10^{-30}</td>
<td>1.340x10^{-26}</td>
<td>53.6</td>
</tr>
</tbody>
</table>

Overall $P$-value for each model is shown. FDR-corrected $P$-values are denoted as $P_{c}$-values. Two regression models, one including clinical factors only and the other including both clinical and genetic factors as covariates, were compared using the likelihood ratio test (LRT).
Table 2.6. Warfarin pharmacogenetic dosing algorithm ($R^2 = 53.6\%$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate</th>
<th>SE</th>
<th>$P$-value</th>
<th>Partial $R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>+6.156</td>
<td>0.743</td>
<td>1.524x10^{-14}</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>-0.029</td>
<td>0.005</td>
<td>1.900x10^{-7}</td>
<td>7.4</td>
</tr>
<tr>
<td>BSA ($m^2$)</td>
<td>+1.114</td>
<td>0.338</td>
<td>0.001</td>
<td>1.8</td>
</tr>
<tr>
<td>VKORC1 -1639 GG</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VKORC1 -1639 GA</td>
<td>-0.632</td>
<td>0.144</td>
<td>0.0119x10^{-3}</td>
<td>5.1</td>
</tr>
<tr>
<td>VKORC1 -1639 AA</td>
<td>-1.627</td>
<td>0.169</td>
<td>2.482x10^{-18}</td>
<td>21.9</td>
</tr>
<tr>
<td>CYP2C9 *1/*1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP2C9 *1/*2</td>
<td>-0.350</td>
<td>0.138</td>
<td>0.012</td>
<td>1.8</td>
</tr>
<tr>
<td>CYP2C9 *1/*3</td>
<td>-0.850</td>
<td>0.169</td>
<td>0.001x10^{-3}</td>
<td>6.8</td>
</tr>
<tr>
<td>CYP2C9 *2/*2, *2/*3</td>
<td>-0.662</td>
<td>0.290</td>
<td>0.023</td>
<td>1.1</td>
</tr>
<tr>
<td>CYP2C9 *3/*3</td>
<td>-3.376</td>
<td>0.512</td>
<td>3.708x10^{-10}</td>
<td>6.8</td>
</tr>
<tr>
<td>Amiodarone status</td>
<td>-0.594</td>
<td>0.185</td>
<td>0.002</td>
<td>2.8</td>
</tr>
</tbody>
</table>

VKORC1 -1639 GG and CYP2C9 *1/*1 were the reference genotypes. The model predicts the square root of weekly dose, therefore the dose in mg/week is obtained by squaring the result. The percentage of warfarin dose variability ($R^2$) explained by each variable is shown. BSA, body surface area; SE, standard error.
2.3.5 Evaluation of the pharmacogenetic algorithm and comparison with published algorithms

To evaluate the performance of our pharmacogenetics algorithm, comparisons were made with five previously published algorithms as follows:

(i) The IWPC pharmacogenetic algorithm which was developed using multi-ethnic populations (Klein et al., 2009).

(ii) The pharmacogenetic algorithm developed in a Central Italian population by Borgiani et al. (2009).

(iii) The pharmacogenetic algorithms developed in Northern Italians by Cini et al. (2012) and Zambon et al. (2011).

(iv) The Couma-Gen algorithm developed in a American-Caucasian population (Anderson et al., 2007).

Table 2.7 details the clinical and genetic factors incorporated in each algorithm and results of their performance accuracy.

Similar to the IWPC pharmacogenetic algorithm which accounted for 54% warfarin dose variance, our pharmacogenetic algorithm explained 53.6% of warfarin dose variability. When using the three Italian pharmacogenetic algorithms significantly lower warfarin dose variance was explained (Zambon et al., 40.9%; Cini et al., 47.1%; Borgiani et al., 31.8%). The Couma-Gen algorithm explained the least variability in warfarin dose (25.6%).

To gain further insight into the performance accuracy of the pharmacogenetic algorithms, the MAE between the observed and predicted doses was calculated. As shown in table 2.7, the IWPC algorithm had the lowest MAE value of 7.19 mg/week. Our algorithm had a MAE of 7.41 mg/week which was very similar to that of IWPC algorithm. The algorithms by Cini et al.,
and Borgiani et al., performed worst with the largest MAE of 10.53 and 10.07 mg/week, respectively.

To evaluate the clinical accuracy of the pharmacogenetic algorithms, we calculated the percentage of patients whose predicted dose fell within 20% of the observed maintenance dose. With our algorithm, the predicted dose of 44.4% of patients fell within 20% of the observed dose and the same percentage was obtained using the IWPC algorithm, suggesting comparable clinical accuracy. The algorithm developed by Cini and colleagues showed the lowest percentage of patients within 20% of the maintenance dose (25.3%).

Positive correlation between the observed and calculated dose was observed with all algorithms (Figure 2.7). The IWPC algorithm achieved the strongest correlation (R=0.712). A fairly strong correlation was observed with our pharmacogenetic algorithm (R=0.687). Poorest correlation was observed with the algorithm by Borgiani and colleagues and the Couma-Gen algorithm.

In conclusion, the performance of the developed algorithm in the studied population from Southern Italy was higher than that achieved when using the other considered algorithms in the studied population, and comparable with the IWPC algorithm.
Table 2.7. Mean absolute errors ± standard errors of weekly warfarin dose and adjusted coefficient of determination in the studied population from Southern Italy obtained using different pharmacogenetic algorithms.

<table>
<thead>
<tr>
<th>Pharmacogenetic algorithm</th>
<th>Country</th>
<th>Ancestry</th>
<th>n</th>
<th>MAE ± SE (mg/wk)</th>
<th>R² adj (%)</th>
<th>Predicted dose within 20% (%)</th>
<th>Clinical parameters</th>
<th>Genetic parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present study algorithm</td>
<td>Southern Italy</td>
<td>Caucasian</td>
<td>266</td>
<td>7.41 ± 0.44</td>
<td>53.6</td>
<td>44.4</td>
<td>Age, BSA, CM</td>
<td>CYP2C9<em>2, CYP2C9</em>3, VKORC1-1639G&gt;A</td>
</tr>
<tr>
<td>Zambon et al., 2011</td>
<td>Northern Italy</td>
<td>Caucasian</td>
<td>274</td>
<td>8.46 ± 0.47</td>
<td>43.6</td>
<td>40.9</td>
<td>Age, BSA</td>
<td>CYP2C9<em>2, CYP2C9</em>3, CYP4F2, 1297G&gt;A, VKORC1-1639G&gt;A</td>
</tr>
<tr>
<td>Cini et al., 2012</td>
<td>Northern Italy</td>
<td>Caucasian</td>
<td>55</td>
<td>10.53 ± 0.53</td>
<td>47.1</td>
<td>25.3</td>
<td>Age, gender, height, weight, smoking status, indication for VTE, diabetes, vegetable intake</td>
<td>CYP2C9<em>2, CYP2C9</em>3, VKORC1-1639G&gt;A, VKORC1 3730G&gt;A</td>
</tr>
<tr>
<td>Borgiani et al., 2009</td>
<td>Central Italy</td>
<td>Caucasian</td>
<td>141</td>
<td>10.07 ± 0.51</td>
<td>31.8</td>
<td>32.8</td>
<td>Age, weight</td>
<td>CYP2C9<em>2, CYP2C9</em>3, CYP4F2, 1297G&gt;A, VKORC1 1173C&gt;T, VKORC1 3730G&gt;A</td>
</tr>
<tr>
<td>Couma-Gen algorithm, 2007</td>
<td>USA</td>
<td>European-American</td>
<td>206</td>
<td>9.96 ± 0.57</td>
<td>25.6</td>
<td>31.1</td>
<td>Age, gender, weight</td>
<td>CYP2C9<em>2, CYP2C9</em>3, VKORC1 1173C&gt;T</td>
</tr>
<tr>
<td>IWPC algorithm, 2009</td>
<td>Various</td>
<td>Caucasian (55.2%), Asian (30.4%), African-American (8.7%), mixed/missing (5.6%)</td>
<td>4043</td>
<td>7.19 ± 0.44</td>
<td>54.0</td>
<td>44.4</td>
<td>Age, height, weight, race, CM</td>
<td>VKORC1-1639G&gt;A, CYP2C9<em>2, CYP2C9</em>3</td>
</tr>
</tbody>
</table>

MAE, mean absolute error; SE, standard error; R² adj, adjusted coefficient of determination; Predicted dose within 20%: percentage of patients whose predicted warfarin dose was within 20% of the actual stable therapeutic dose; BSA, body surface area; CM, concomitant medications; VTE, venous thromboembolism.
Figure 2.7. Correlation between observed and predicted doses. The correlations between the observed weekly stable dose and the predicted dose calculated using the pharmacogenetic algorithm from (A) our study, (B) Zambon et al. (2011), (C) Cini et al. (2012), (D) Borgiani et al. (2009), (E) Anderson et al. (2007), (F) Klein et al. (2009) are shown. The Pearson correlation coefficient (R) and the P-value are shown in the upper left corner.
2.4 DISCUSSION

In agreement with published literature, homozygous carriers of the minor allele for CYP2C9 and/or VKORC1 -1639G>A in our cohort of patients were associated with lower warfarin dose requirement with respect to heterozygotes and carriers with homozygous wild-type genotype. Conversely, for the VKORC1 3730G>A polymorphism which is known to be associated with warfarin resistance (Geisen et al., 2005; Rieder et al., 2007), homozygous AA patients required significantly higher warfarin dose in comparison to heterozygous and homozygous GG patients.

Although VKORC1 3730G>A polymorphism was found significant in the univariate analysis, it failed to be included in the algorithm, suggesting that this SNP is probably not predictive in our Southern Italian population. Other Italian studies with smaller cohorts of patients from the Northern and Central regions of Italy demonstrated that VKORC1 3730G>A contributes to 10% warfarin dose variance (Borgiani et al., 2009; Cini et al., 2012). However, further studies in larger Italian cohorts are needed to confirm this finding.

Borgiani and colleagues (2007) have reported that the VKORC1 1173C>T polymorphism account for 20% of the total mean weekly warfarin dose in their cohort of patients from Central Italy. Studies have shown that the variant allele of VKORC1 1173C>T is associated with reduced warfarin dose requirement (D’Andrea et al., 2005; Borgiani et al., 2007), but this SNP was not associated with warfarin stable dose in our study. Taking a closer look, the minor allele frequency of VKORC1 1173C>T appears to be lower in Southern Italians (13.5%, our study; 16.3%, D’Andrea et al., 2005) when compared to
Italians from the central region (26%, Borgiani et al., 2007). This observed genetic difference suggests intra-ethnic heterogeneity between Southern and Central Italians. Together with different cultural and lifestyle factors, this could in part explain discrepancy observed with warfarin dose association. In addition, several studies have shown that the \textit{VKORC1} 1173C>T SNP is associated with warfarin sensitivity in Caucasians (Carlquist et al., 2006; Veenstra et al., 2005; Wadelius et al., 2005). However, this was not observed in our population as no sensitive patients requiring extreme low doses were present in our cohort of patients.

Reports have shown that carriers of the \textit{CYP4F2} rs2108622 variant T allele require approximately 1-2.5 mg more warfarin per day than those carrying the wild-type C allele (Borgiani et al., 2009; Caldwell et al., 2008; Takeuchi et al., 2009). Although studies conducted in patients from Northern and Central Italy have found significant association between \textit{CYP4F2} rs2108622 and warfarin stable dose (Borgiani et al., 2009; Zambon et al., 2011), no association was found in our cohort of patients from Southern Italy. Furthermore, no association was observed in African-Americans (Cavallari et al., 2010) or Brazilians (Perini et al., 2010). The minor allele frequency of \textit{CYP4F2} rs2108622 appears to be lower in our population (8.3%) when compared to Italians from the Northern (9.4%, Zambon et al., 2011) and Central Italy (12%, Borgiani et al., 2009), suggesting again inter- and intra-ethnic differences across the Country.

Several pharmacogenetic algorithms have previously been developed with the aim to improve the accuracy of warfarin dose prediction (Anderson et al., 2007; Borgiani et al., 2009; Cini et al., 2012; Klein et al., 2009; Schelleman
et al., 2008; Verhoef et al., 2014; Zambon et al., 2011). However, each algorithm may be biased when applied to a population other than that from which it is derived, mainly because of the genetic differences between the studied populations. For instance, the genetics of American-Caucasians may differ from the Europeans (Anderson et al., 2007). Furthermore, within the Europeans, the genetics of Northern Europeans may differ from those of Southern Europeans (Schelleman et al., 2008). In Italy, genetic heterogeneity has been observed between Northern, Central and Southern Italians (Limdi et al., 2008a; Schelleman et al., 2008).

In this chapter, we have developed the first pharmacogenetic warfarin dosing algorithm in a Southern Italian population. Our pharmacogenetic algorithm incorporated six variables (age, BSA, amiodarone status, \( \text{CYP2C9}^*2 \), \( \text{CYP2C9}^*3 \), and \( \text{VKORC1} \ -1639G>A \)) and allowed accurate prediction of warfarin maintenance dose in 44% of patients in our cohort, with a mean absolute error of 7.41 mg/week. The prediction accuracy of our pharmacogenetic algorithm was similar to that of the IWPC algorithm. Given that the IWPC algorithm was derived from a large admixed population, genetic differences across different ethnic groups were well captured, allowing high prediction accuracy of warfarin dose.

Interestingly, our pharmacogenetic algorithm has demonstrated superior prediction accuracy to the three published Italian pharmacogenetic algorithms derived from patients in Northern and Central Italy (Borgiani et al., 2009; Cini et al., 2012; Zambon et al., 2011). Indeed, genetically homogeneous populations do not always coincide with the ‘political’ definition of a country. The genetic differences observed across Italy could be a
consequence of demographic processes such as internal migration within and between the macro-areas. Indeed, Italy remains characterized by a strong migratory movement of the population within its territory (Bonifazi and Heins, 2000) which was particularly significant between 1959–1970, but still present nowadays. Within Italy, allele frequency differences warrant caution when matching patients from different areas of the Country. Instead of implementing a universal pharmacogenetic dosing algorithm for the Italian population, tailored algorithms for the Northern, Central and Southern regions of Italy may be necessary to increase the accuracy of warfarin dose prediction.

Whilst our pharmacogenetic algorithm has accounted for more than 50% of warfarin dose variability, over 40% of warfarin dose variance still remains unexplained. Moreover, our pharmacogenetic algorithm does not predict accurately for patients who require low- or high-dose regimens, which is a common observation in previously published algorithms (Anderson et al., 2007; Klein et al., 2009). Further exploration in transcriptomics and rare variants may help elucidate the unexplained dose variability.
CHAPTER 3

Circulating miRNAs as novel biomarkers of response to warfarin
CHAPTER 3

CONTENTS

3.1 INTRODUCTION ............................................................... 93

3.2 MATERIALS AND METHODS .............................................. 97
  3.2.1 Subjects ..................................................................... 97
  3.2.2 Samples and miRNA isolation from serum .................. 98
  3.2.3 miRNA arrays .......................................................... 101
  3.2.4 Haemolysis evaluation .............................................. 103
  3.2.5 Real-time qPCR standard curve optimisation – Phase 1 .... 103
    3.2.5.1 miRNA reverse transcription ................................. 103
    3.2.5.2 Pre-Amplification ............................................... 107
    3.2.5.3 Real-time qPCR .................................................. 107
  3.2.6 Real-time qPCR standard curve optimisation – Phase 2 .... 110
    3.2.6.1 miRNA reverse transcription ................................. 110
    3.2.6.2 Pre-Amplification ............................................... 110
    3.2.6.3 Real-time qPCR .................................................. 110
  3.2.7 Real-time qPCR standard curve optimisation – Phase 3 .... 112
    3.2.7.1 miRNA reverse transcription ................................. 112
    3.2.7.2 Pre-Amplification ............................................... 112
    3.2.7.3 Real-time qPCR .................................................. 112
  3.2.8 Validation procedure .................................................. 114
  3.2.9 Statistical analysis .................................................... 115

3.3 RESULTS .............................................................................. 116
  3.3.1 miRNA array quality control and normalisation .......... 116
  3.3.2 Patient characteristics .............................................. 118
  3.3.3 Haemolysis evaluation .............................................. 121
  3.3.4 miRNA association analysis ....................................... 122
    3.3.4.1 Bioinformatics analysis ....................................... 125
    3.3.4.2 Identification of endogenous controls .................... 128
  3.3.5 miRNA validation ...................................................... 128
  3.3.6 Real-time PCR standard curve optimisation .................. 130
    3.3.6.1 Phase 1 ............................................................ 130
    3.3.6.2 Phase 2 ............................................................ 132
CHAPTER 3

3.3.6.3 Phase 3 ......................................................................................... 135

3.4 DISCUSSION ....................................................................................... 138
3.1 INTRODUCTION

The predictive power of pharmacogenetic-based dosing algorithms, encompassing genetic and non-genetic covariates, accounts for no more than 60% of the variance in warfarin dose requirement (Klein et al., 2009; Verhoef et al., 2014; Wadelius et al., 2009). 30-40% of the variance in daily warfarin dose requirement remains unexplained, suggesting that other factors might be important.

Since the discovery of the presence of extracellular microRNAs (miRNAs) in a variety of bodily fluids, a large number of studies have investigated the potential use of cell-free miRNAs as diagnostic and/or prognostic biomarkers for various diseases (Bertoli et al., 2015; Churov et al., 2015; Creemers et al., 2012; Hung et al., 2015; Mo et al., 2012; Reid et al., 2011). Being easily accessible and collected routinely during medical assessments, plasma and serum represent the most promising and best-studied source of cell-free miRNAs.

Using bioinformatic tools, Shomron (2010) reported two highly evolutionary conserved binding sites for miR-133a and miR-137 on the VKORC1 gene, suggesting that VKORC1 could be regulated by these miRNAs. Interestingly, Pérez-Andreu et al. (2012) showed that miR-133a is constitutively co-expressed with VKORC1 in human hepatocytes. The authors also demonstrated that the over-expression of miR-133a resulted in decreased VKORC1 mRNA expression in a dose dependent-manner. Moreover, subjects carrying the G allele for the VKORC1 promoter SNP (rs9923231) showed association with higher VKORC1 mRNA transcription rate.
miR-133a was first experimentally characterised in mice (Lagos-Quintana et al., 2002). In the human genome, there are three known miR-133 genes: MIR133A1, MIR133A2, and MIR133B found on chromosomes 18q11.2, 20q13.33 and 6p12.2, respectively. MIR133A1 and MIR133A2 encode miR-133a-1 and miR-133a-2, which present the same identical mature sequence. MIR133B codes for miR-133b, which differs from the others by a single nucleotide at the 3' end (see Table 3.1). Both miR-133a-1 and miR-133a-2 are canonically expressed in skeletal and cardiac muscles, whereas miR-133b is expressed in developing skeletal muscles but not in cardiac muscles (Liu et al., 2008b; Townley-Tilson et al., 2010).

**Table 3.1. miR-133 family.**

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Mature sequence</th>
<th>Gene</th>
<th>Chr</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-133a-1</td>
<td>UUUGGUCCCUCUCAACCAGCUG</td>
<td>MIR133A1</td>
<td>18</td>
</tr>
<tr>
<td>miR-133a-2</td>
<td>UUUGGUCCCUCUCAACCAGCUG</td>
<td>MIR133A2</td>
<td>20</td>
</tr>
<tr>
<td>miR-133b</td>
<td>UUUGGUCCCUCUCAACCAGCUA</td>
<td>MIR133B</td>
<td>6</td>
</tr>
</tbody>
</table>

miR-133a is one of the most studied and best characterized miRNAs to date (Mitchelson and Qin, 2015). Abundantly expressed in myocardial cells (Chen et al., 2006; Fichtlscherer et al., 2010; Wang et al., 2010), miR-133a has been demonstrated to play an important role in myogenesis, cardiac development and hypertrophy (Carè et al., 2007; He et al., 2011; Li et al., 2010b; Liu et al., 2011a, 2008b). Previous studies have demonstrated that miR-133a is present at low levels in the plasma of healthy individuals, and it is differentially expressed in different cardiovascular diseases (D’Alessandra et
Increased circulating miR-133a levels were observed in the early phase of acute myocardial infarction and in patients with coronary artery stenosis, suggesting a possible role for this miRNA in the diagnosis of coronary heart disease (Wang et al., 2010).

Beyond the study in muscles, miR-133a has been implicated in cancer and identified as a key factor in cancer development and progression, (Dong et al., 2013; Dyrskjøt et al., 2009; Gutiérrez et al., 2010; Marcucci et al., 2010; Qin et al., 2012; Wan et al., 2014). Much attention has also been drawn to the versatile molecular functions of miR-133a, making it a truly valuable therapeutic gene in miRNA-based gene therapy (Kano et al., 2010).

Recently, Ciccacci and colleagues (2015) performed direct sequencing of the MIR133A1, MIR133A2, and MIR133B genes in 205 patients treated with warfarin, and found that polymorphisms in miR-133 genes influence warfarin dose requirements. Interestingly, patients carrying the variant A allele for the MIR133A2 SNP (rs45547937) required significantly higher mean weekly warfarin dose than subjects with wild-type genotype (GG) (28.17 mg/week versus 22.39 mg/week respectively, P=0.019) (Ciccacci et al., 2015). In the linear regression model built by the authors, this SNP accounted for 2.3% of warfarin dose variability, an effect much less than that observed with polymorphisms in VKORC1 and CYP2C9 genes (Ciccacci et al., 2015).

Taken together, data from Shomron (2010), Pérez-Andreu et al. (2012) and Ciccacci et al. (2015) suggest that miR-133a may play a role in warfarin response. Given that the role of circulating miRNA in warfarin response has never been investigated, the aim of this chapter was therefore to evaluate the potential effects of miR-133a and other miRNAs across the
human transcriptome in serum from warfarinised patients, using a miRNA array. In addition, the TaqMan real-time quantitative PCR (qPCR) procedure was optimised to assess and validate miRNA expression levels in patients’ sera.
3.2 MATERIALS AND METHODS

3.2.1 Subjects

Between November 2004 and May 2007, patients ($n=1000$) starting warfarin therapy were prospectively recruited from two hospitals in Liverpool, the Royal Liverpool University Hospital and University Hospital Aintree. The study was approved by the Birmingham South research ethics committee and written informed consent was obtained from all patients (Jorgensen et al., 2009; Zhang et al., 2009).

Stable warfarin dose was defined as an unchanged daily dose at three or more consecutive clinic within the individual’s target INR range. A bleeding event was defined as major or minor according to the classification by Fihn and colleagues (1996). A bleeding event was classed as serious if it was lethal, life-threatening, permanently disabling or led to hospital admission or prolongation of hospital stay. Only haemorrhagic events considered to be possibly, probably or definitely associated with warfarin were included in the analyses.

For the pilot study, 12 subjects who have achieved warfarin stable dose and 11 subjects who did not achieve warfarin stable dose were selected from this prospective cohort of patients. Given that clinical and genetic factors such as age, BMI, $CYP2C9^*$2, $CYP2C9^*$3 and $VKORC1$ genotypes have been shown to affect warfarin dose requirements, all the patients were matched for these factors.
3.2.2 Samples and miRNA isolation from serum

Venous blood was collected in a blood tube containing clot activator from all the patients. The blood was processed for the isolation of serum by spinning at 2,600 x g for 20 min at room temperature. Serum was aliquoted into sterile cryovials and stored at -80°C for later use.

Total RNA, including miRNA was isolated from 200 µL of serum using Norgen’s Total RNA Purification kit (Product #37500, Norgen Biotek Corporation, Canada), according to the manufacturer’s instructions for serum samples. Frozen serum aliquots were defrosted and centrifuged at 21,100 x g for 10 min at RT. The clear supernatant was used for downstream extraction. The Norgen’s Purification Kit, which utilizes a silica-based nucleic acid purification approach with no organic compounds provides a fast, reliable reproducible and simple procedure for isolating circulating RNA and exosomal RNA from small serum input without the use of phenol or chloroform. Purification is based on Norgen’s new spin column chromatography that uses Norgen’s proprietary resin separation matrix. The kit is designed to isolate all sizes of circulating RNA, including microRNA, as well as all sizes of exosomal RNA from plasma or serum samples. The summary of the workflow to purify total RNA from serum using Norgen’s Total RNA purification kit is shown in Figure 3.1.

The amount of miRNA present in serum may be influenced either by miRNA extraction efficiency or the robustness of the quantification method. In order to adjust for such variations, the synthetic Caenorhabditis Elegans miR-39 (cel-miR-39-3p: 5’-UCACCGGGUGUAAUCACGUUG-3’, Qiagen) was added as a spike-in control to check for extraction efficiency. 3.5 µL of the
synthetic cel-miR-39 (1.6x10^8 copies/µL) was added after the lysis step and went through the entire RNA isolation process, acting as an internal reference for normalization (exogenous control) of technical variations between samples.

To remove any residual DNA that may affect downstream applications, an On-Column DNA Removal step was performed using Norgen’s RNase-free DNase I Kit (Product #25710, Norgen Biotek Corporation, Canada). Each sample was eluted in 50 µL of Elution Solution, provided with the kit, and aliquoted into two RNase-free tubes.

RNA concentration and purity were evaluated using NanoDrop 8000 spectrophotometer (Thermo Scientific). The A260/230 and A260/280 ratios were used to assess the presence of contaminants. Total RNA was then stored at -80°C.
Figure 3.1. Flowchart. Summary of workflow for purifying total RNA from serum using Norgen’s Total RNA purification kit (Product #37500, Norgen Biotek Corporation, Canada).


3.2.3 miRNA arrays

To investigate the miRNA profile across the whole human transcriptome, 150 ng of total RNA extracted from each of the 23 patients’ sera was sent to the service provider, Edinburgh Genomics (Roslin Institute, UK), to be processed on the Affymetrix GeneChip® miRNA 4.1 Array plate (Affymetrix, Santa Clara, CA, USA). The 23 samples were assayed in a 24-array plate together with one internal control.

The Affymetrix array provides 100% miRBase v20 coverage (www.mirbase.org) by a one-color approach (Liu et al., 2004, 2008a). The microarray detects 2,578 human mature miRNA, 2,025 human pre-miRNA, and 1,996 human snoRNA, CDBox RNA, H/ACA Box RNA and scaRNA, and provides 4-log dynamic range, with >95% reproducibility and 85% transcript detection at 1.0 amol for a total RNA input of 130-500ng. Figure 3.2 describes the workflow of an Affymetrix GeneChip experiment.

The Affymetrix® Expression Console™ Software was used to perform data quality control and normalisation. Statistical tests for differential expression were performed using the t-test in the MultiExperiment Viewer (MeV) v4.9 with 180,000 permutations and a fold change threshold >1.5. Bioinformatics analyses to investigate the biological functions of miRNAs’ target genes were carried out using the freely available online software, Database for Annotation, Visualization and Integrated Discovery (DAVID v6.7 (Huang et al., 2009a, 2009b).
Figure 3.2. Schematic overview on the different steps while performing an Affymetrix GeneChip experiment. Starting from total RNA, the first step involves a brief tailing reaction followed by ligation of the biotinylated signal molecule to the target RNA sample, using FlashTag™ Biotin HSR Labelling Kit (Affymetrix, Santa Clara, CA, USA). The high sensitivity of FlashTag Biotin HSR is due to proprietary 3DNA® dendrimer signal amplification technology. The 3DNA dendrimer is a branched structure of single and double stranded DNA conjugated with numerous labels, which provides ultrasensitive biotin labelling (Nilsen et al., 1997; Stears et al., 2000). The Biotin-labelled RNA is then added to the array. Where a miRNA fragment and a probe are complimentary, the miRNA hybridizes to the probes in the feature. The array is then washed to remove any unstuck miRNA (i.e., no match was made) and then stained with the fluorescent molecule that sticks to Biotin (streptavidin-phycoerythrin). Lastly, the entire array is scanned with a laser and the information is kept in a computer for quantitative analysis. Adapted from Affymetrix® (www.affymetrix.com).
3.2.4 Haemolysis evaluation

Serum absorbance levels were measured spectrophotometrically using the DTX 880 Multimode Detector (Beckman Coulter®, UK). The haemoglobin concentration in each diluted sample (1:10 in water) was calculated using the Harboe method (1959) with the following formula:

\[
\text{Haemoglobin (g/L)} = (167.2 \times A_{414} - 83.6 \times A_{380} - 83.6 \times A_{450}) \times 10/1000
\]

where \( A_{414} \), \( A_{380} \), and \( A_{450} \) are the absorption at 414, 380, and 450 nm respectively, 10 is the dilution factor, and 1000 is the conversion factor. This method uses the known absorption coefficients of haemoglobin and bilirubin at 414 and 450 nm. Samples with haemoglobin values \( \geq 0.1 \) g/L were considered haemolysed.

3.2.5 Real-time qPCR standard curve optimisation – Phase 1

3.2.5.1 miRNA reverse transcription

Standard curves were generated using chemically synthesised RNA oligonucleotides corresponding to the 5 miRNAs of interest. Optimisations were carried out to test for assay efficiency, precision, sensitivity, and working range. The synthetic miRNAs were purchased from Integrated DNA Technologies (IDT, USA). Sequences of the synthetic miRNAs are listed in Table 3.2. The synthetic miRNAs of interest were serially diluted (1:10) from a
1 mM stock to a final concentration of 1000, 100, 10, 1, and 0.1 pM in RNase-free water. The optimisation workflow is shown in Figure 3.3.

Input miRNA was reverse-transcribed using TaqMan MicroRNA Reverse Transcription Kit and custom Multiplex RT Primer pool (Thermo Fisher Scientific) in a Veriti Thermal Cycler (Applied Biosystems). A fixed volume of 2 µL miRNA was used as template in the RT reaction for each dilution of the miRNAs of interest, and combined with 3 µL of an equimolar (0.05X) mix of custom Multiplex RT primer pool, 2 mM dNTPs (with dTTP), 1X Reverse Transcription Buffer, 0.25 U/µL RNase Inhibitor, 10 U/µL MultiScribe Reverse Transcriptase, and RNase-free water in a final reaction volume of 7.5 µL. The multiplex RT primer pool consisted of primers for miR-4487, -548a, -133a-3p, -92a-3p, and cel-miR-39-3p (Thermo Fisher Scientific). No-template samples were included as negative controls.

Reaction conditions followed the manufacturer’s instructions: annealing for 30 min at 16°C, followed by complementary DNA (cDNA) synthesis for 30 min at 42°C and denaturation for 5 min at 85°C. The resulting cDNA was diluted 1:3 with RNase-free water and stored at -20°C.
<table>
<thead>
<tr>
<th>miRNA</th>
<th>miRBase Accession Number</th>
<th>miRBase Version</th>
<th>Species</th>
<th>Chr</th>
<th>Mature sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>cel-miR-39-3p</td>
<td>MI0000010</td>
<td>21</td>
<td><em>Caenorhabditis Elegans</em></td>
<td>-</td>
<td>5'-UCACCGGGUGUAAUUCAGCUUG-3'</td>
</tr>
<tr>
<td>hsa-miR-4487</td>
<td>MIMAT0019021</td>
<td>21</td>
<td>Human</td>
<td>11</td>
<td>5'-AGAGCUGGCUAGGGCAG-3'</td>
</tr>
<tr>
<td>hsa-miR-92a-3p</td>
<td>MI0000093</td>
<td>21</td>
<td>Human</td>
<td>13</td>
<td>5'-UAUUGCACUUGUCCGGCCUGU-3'</td>
</tr>
<tr>
<td>hsa-miR-133a-3p</td>
<td>MI0000450</td>
<td>21</td>
<td>Human</td>
<td>18</td>
<td>5'-UUUGGUCCCCUCAACGACUG-3'</td>
</tr>
<tr>
<td>hsa-miR-548a-3p</td>
<td>MI0003593</td>
<td>21</td>
<td>Human</td>
<td>6</td>
<td>5'-CAAAACUGGCAUUACUUUUGC-3'</td>
</tr>
</tbody>
</table>
Figure 3.3. Workflow of phase 1 standard curve optimisation. Schematic representation of serial dilutions of synthetic miRNAs. The 5 serial dilutions were reverse transcribed and pre-amplified. Pre-amplification products were diluted 1:160 in 0.005XTE before the real-time qPCR assay.
3.2.5.2 Pre-Amplification

Pre-amplification of cDNA was run on a Veriti Thermal Cycler (Applied Biosystems). 1.5 µL of diluted (1:3) cDNA was combined with 0.75 µL of custom Pre-Amplification Primer pool and 2.75µL of 2X TaqMan PreAmp Master Mix (Thermo Fisher Scientific) to make a final reaction volume of 5 µL. Pre-Amplification Primer pool was prepared so that the final concentration of each primer equalled 0.2X. Negative controls containing RNase-free water instead of cDNA were included. Reaction conditions followed the manufacturer’s instructions: 95°C for 10 min, 55°C for 2 min, 72°C for 2 min, followed by 12 cycles at 95°C for 15 sec and 60°C for 4 min. The reaction was terminated with a 10 min final enzyme inactivation at 99.9°C. Pre-amplification products were stored at -20°C.

3.2.5.3 Real-time qPCR

TaqMan® miRNA Assays (Thermo Fisher Scientific) for the 5 specific miRNAs of interest, cel-miR-39-3p, hsa-miR-4487, hsa-miR-548a, hsa-miR-133a-3p, and hsa-miR -92a-3p, were used to perform quantitative PCR (qPCR). The chemistry of real-time qPCR detection of miRNAs is summarised in Figure 3.4.

The pre-amplification products were diluted 1:160 in 0.005XTE, and 2 µL was used as a template in a total volume of 10 µL containing 1X TaqMan microRNA assay, and 1X TaqMan Universal Master Mix II, no AmpErase UNG. No-template samples were included as negative controls.
The reaction was performed on an ABI 7900HT Fast Real-Time PCR System (Applied Biosystems) at the following conditions: 10 min enzyme activation at 95°C followed by 40 cycles of 15 sec denaturation at 95°C and 60 sec annealing/elongation at 60°C.

Data were analysed using SDS Software version 2.3 (Applied Biosystems). Cycle threshold (Ct) values were determined by applying a fixed fluorescence threshold level of 0.04 across all miRNAs. miRNAs with detectable transcription were defined as those with a Ct less than 35. The Ct is defined as the number of cycles required for the fluorescent signal to cross the threshold in real-time PCR. Plotting the Ct values versus the logarithmic concentration of the synthetic miRNA in a standard curve allows for fitting a curve to determine the miRNA concentration from the Ct values of the biological samples.
**Figure 3.4. TaqMan® MicroRNA Assay.** miRNA quantification by TaqMan® MicroRNA assay is done using two-step RT-PCR. In the reverse transcription (RT) step, cDNA is reverse transcribed from total RNA using specific miRNA primers. The stem–loop structure provides specificity for only the mature miRNA target and forms a RT primer/mature miRNA chimera that extends the 3’ end of the miRNA. The resulting longer RT-product presents a template amenable to standard real-time PCR using TaqMan® Assays. In this study, RT-product was pre-amplified before the PCR step to increase the amount of starting material. The TaqMan MGB probes contains: (i) a reporter dye (FAM™ dye) linked to the 5’ end of the probe; (ii) a minor groove binder (MGB) at the 3’ end of the probe, to increase the melting temperature without increasing probe length; and (iii) a non-fluorescent quencher (NFQ) at the 3’ end of the probe. During PCR, the TaqMan MGB probe anneals specifically to the complementary sequence between the forward and reverse primer sites. When the probe is intact, the proximity of the reporter dye to the quencher dye results in suppression of the reporter fluorescence. The DNA polymerase cleaves only probes that are hybridized to the target. Cleavage separates the reporter dye from the quencher dye, resulting in increased fluorescence by the reporter. The increase in fluorescence signal occurs only if the target sequence is complementary to the probe and is amplified during PCR; therefore, nonspecific amplification is not detected.
3.2.6 Real-time qPCR standard curve optimisation – Phase 2

3.2.6.1 miRNA reverse transcription

The synthetic miRNAs of interest were serially diluted (1:10) from a 1 mM stock to a final concentration of 100, 10, 1, 0.1, and 0.01 µM in RNase-free water. The workflow of phase 2 standard curve optimisation is shown in Figure 3.5. Input miRNA was reverse-transcribed as described in section 3.2.5.1.

3.2.6.2 Pre-Amplification

Pre-amplification conditions are described in section 3.2.5.2.

3.2.6.3 Real-time qPCR

The pre-amplified synthetic miRNAs were serially diluted (1:10) five times using 0.1XTE. The diluted pre-amplified products were further diluted 1:160 using 0.005XTE, and 2 µL were used as a template as previously described (see section 3.2.5.3). No-template samples were included as negative controls. To reduce wastage of reagents, phase 2 optimisation was performed using only two miRNAs, miR-133a-3p and miR-548a-3p.
Figure 3.5. Workflow of phase 2 standard curve optimisation. Schematic representation of serial dilutions of synthetic miRNAs. Each of the serially diluted synthetic miRNAs was reverse transcribed and pre-amplified. Five serial dilutions (1:10) of the pre-amplification products were made using 0.1XTE. The diluted pre-amplification products were further diluted 1:160 in 0.005XTE before the real-time qPCR assay.
3.2.7 Real-time qPCR standard curve optimisation – Phase 3

3.2.7.1 miRNA reverse transcription

The synthetic miRNAs of interest were diluted from a 1 mM stock to a final concentration of 1 µM in RNase-free water and reverse-transcribed as described in section 3.2.5.1. The workflow of phase 3 standard curve optimisation is shown in Figure 3.6.

3.2.7.2 Pre-Amplification

Pre-amplification conditions are described in section 3.2.5.2.

3.2.7.3 Real-time qPCR

To capture the Ct range between 5 and 35, an eight-point standard curve, ranging from 1 µM to 100 fM, was created. Seven 1:10 serial dilutions from the 1 µM pre-amplified synthetic miRNAs were carried out using 0.1XTE. All the diluted synthetic miRNAs were further diluted 1:160 in 0.005XTE, and 2 µL was assayed as previously described (see section 3.2.5.3). No-template samples were included as negative controls. All reactions, including the negative controls, were performed in triplicate.
Figure 3.6. Workflow of phase 3 standard curve optimisation. Schematic representation of serial dilutions of synthetic miRNAs. The synthetic miRNAs of interest were diluted from a 1 mM stock to a final concentration of 1 µM in RNase-free water, reverse transcribed and pre-amplified. Seven serial dilutions (1:10) of pre-amplification products were made using 0.1XTE. The diluted pre-amplification products were further diluted 1:160 in 0.005XTE before carrying out the real-time qPCR assay.
3.2.8 Validation procedure

Three hallmarks were considered when optimising the standard curves during qPCR:

i. Linear standard curve $R^2 > 0.98$

ii. Consistency across replicates

iii. High amplification efficiency (90-105%)

The correlation coefficient ($R^2$) is a measure of how well the data fits the model and how well the data fits on a straight line, which is influenced by pipetting accuracy and by the sensitivity range of the assay. If $R^2$ is ≤0.98, the assay may not give reliable results.

Consistency across replicates to determine the accuracy and the precision of the method was evaluated by the coefficient of variation (CV). CV is a standardized measure of dispersion of a probability distribution or frequency distribution, and is defined as the ratio of the standard deviation to the mean.

PCR amplification efficiency is the rate at which a PCR amplicon is generated, commonly expressed as a percentage value. If a particular PCR amplicon doubles in quantity during the geometric phase of its PCR amplification, then the PCR assay has 100% efficiency.

The slope of the standard curve was used to estimate the PCR amplification efficiency of every real-time PCR reaction. The real-time PCR standard curve was graphically represented as a semi-logarithmic regression line plot of Ct value versus logarithm of input RNA. A standard curve slope of -3.32 indicates a PCR reaction with 100% efficiency. A 100% efficient reaction yields a 10-fold increase in PCR amplicon every 3.32 cycles during the
exponential phase of amplification \( \log_2 10 = 3.3219 \). Slopes more negative than -3.32 indicate reactions less than 100% efficient. Slopes more positive than -3.32 may indicate reduced sample quality or pipetting problems.

The following formula was used for estimating the efficiency \( (E) \) of every real-time PCR assay:

\[
E = 10^{-1/slope}
\]

Then \( E \) was converted to percentage by:

\[
\% \text{ Efficiency} = (E-1) \times 100
\]

Values between 90% and 105% indicates high amplification efficiency. A slope between -3.9 and -3.0 (80-110% efficiency) is generally acceptable.

### 3.2.9 Statistical analysis

Differences in clinical, amnestic and demographic features among patients were tested by the Student t test and ANOVA for continuous variables, or by Pearson's chi-square test for categorical variables. A \( P<0.05 \) was considered statistically significant. Statistical analysis was performed with SPSS v22.0 (SPSS Inc., Chicago, IL, USA).
3.3 RESULTS

3.3.1 miRNA array quality control and normalisation

The 23 samples were assayed together with one Affymetrix internal control. All the data are shown as relative intensity to the internal control (logarithmic transformed) in Figure 3.7. After quality control by Affymetrix Expression Console, v1.4.1, three outliers were found and excluded from further analysis. Of the remaining 20 samples, 10 have achieved warfarin stable dose while the other 10 did not achieve warfarin stable dose.
Figure 3.7. Signal box plot showing the ratio of individual probe intensity to the median probe intensity across all arrays. The Relative Log Probe Cell intensity values are computed by calculating for each probe-set the ratio between the expression of a probe-set and the median expression of this probe-set across all arrays of the experiment. It is assumed that most probe-sets are not changed across the arrays, so it is expected that these ratios are around 0 on a log scale. The box plots presenting the distribution of these log-ratios should then be centred near 0 and have similar spread. The first boxplot on the left represents the Affymetrix internal control. Outliers are indicated with a red arrow.
3.3.2 Patient characteristics

The clinical, anamnestic and demographic features of the 20 patients are shown in Table 3.3. Both the stable dose and non-stable dose patient groups were matched for age, BMI, and gender. Mean age was 77 years, mean BMI was between 27-28, and the male to female ratio was approximately 2:3. The main indication for warfarin therapy was atrial fibrillation (75.0%). All adverse events which occurred during the study period were reported and assessed for causality. Information on concomitant medication, co-morbidities, as well as details of smoking history, current medications and alcohol intake were gathered from all patients. No significant differences in these variables were found between the groups of stable and non-stable dose patients.
Table 3.3. Clinical, amnestic and demographic features of the warfarin-treated patients (n=20).

<table>
<thead>
<tr>
<th></th>
<th>All (20)</th>
<th>Stable dose (10)</th>
<th>Non-stable dose (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>77.10±3.91</td>
<td>77.13±4.92</td>
<td>77.08±2.84</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>7/13</td>
<td>4/6</td>
<td>3/7</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>27.96±6.41</td>
<td>28.71±6.34</td>
<td>27.20±6.72</td>
</tr>
<tr>
<td>BSA (m²)</td>
<td>1.86±0.21</td>
<td>1.90±0.16</td>
<td>1.81±0.25</td>
</tr>
<tr>
<td>Smoking (yes/no)</td>
<td>5/15</td>
<td>1/9</td>
<td>4/6</td>
</tr>
<tr>
<td>Alcohol intake (units)</td>
<td>3.30±3.25</td>
<td>3.20±3.42</td>
<td>3.40±3.24</td>
</tr>
<tr>
<td>Indication for warfarin therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>15</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Deep venous thrombosis</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>CVA/TIA</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Stable warfarin dose (mg/day)</td>
<td>3.93±1.72</td>
<td>3.93±1.72</td>
<td>-</td>
</tr>
<tr>
<td>All bleeding complications (yes/no)</td>
<td>7/13</td>
<td>2/8</td>
<td>5/5</td>
</tr>
<tr>
<td>Major bleeding complications (yes/no)</td>
<td>2/18</td>
<td>1/9</td>
<td>1/9</td>
</tr>
</tbody>
</table>
### Table 3.3. Continued

<table>
<thead>
<tr>
<th>Comorbidities</th>
<th>All (20)</th>
<th>Stable dose (10)</th>
<th>Non-stable dose (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular disease (yes/no)</td>
<td>2/18</td>
<td>2/8</td>
<td>0/10</td>
</tr>
<tr>
<td>Respiratory disease (yes/no)</td>
<td>4/16</td>
<td>0/10</td>
<td>4/6</td>
</tr>
<tr>
<td>Hepatic disease (yes/no)</td>
<td>2/18</td>
<td>1/9</td>
<td>1/9</td>
</tr>
<tr>
<td>Gastrointestinal disorder (yes/no)</td>
<td>7/13</td>
<td>3/7</td>
<td>4/6</td>
</tr>
<tr>
<td>Urological disease (yes/no)</td>
<td>1/19</td>
<td>1/9</td>
<td>0/10</td>
</tr>
<tr>
<td>Neurological disorder (yes/no)</td>
<td>6/14</td>
<td>2/8</td>
<td>4/6</td>
</tr>
<tr>
<td>Muscular problems (yes/no)</td>
<td>3/17</td>
<td>3/7</td>
<td>0/10</td>
</tr>
<tr>
<td>History of fall (yes/no)</td>
<td>3/17</td>
<td>0/10</td>
<td>3/7</td>
</tr>
<tr>
<td>Endocrine disorders (yes/no)</td>
<td>3/17</td>
<td>1/9</td>
<td>2/8</td>
</tr>
<tr>
<td>Concomitant drug assumption</td>
<td>2/18</td>
<td>2/8</td>
<td>0/10</td>
</tr>
</tbody>
</table>

Data are expressed as average ± SD (continuous variables) or in number of patients (categorical variables).

Stable dose, patients who have achieved warfarin stable dose (n=10); non-stable dose, patients who did not achieve warfarin stable (n=10); CVA, cardiovascular accident; TIA, transient ischemic attack; concomitant drug assumption, one patient on simvastatin and one on ezetimibe during stable dose period.
3.3.3 Haemolysis evaluation

No differences in the sera haemoglobin concentration was observed in patients who achieved warfarin stable dose and patients who did not achieve warfarin stability ($P = 0.192$) (Figure 3.8). All the samples assayed were not haemolysed (haemoglobin values <0.1 g/L).

![Figure 3.8. Haemolysis evaluation in patients’ sera.](image)

*Figure 3.8. Haemolysis evaluation in patients’ sera.* The haemoglobin concentration (g/L) in each diluted serum sample (1:10 in water) was calculated using the Harboe method (1959). $P$-value is shown in the upper right corner. Each dot represents an individual and solid lines represent the mean values. Stable dose, patients who have achieved warfarin stable dose (n=10); non-stable dose, patients who did not achieve warfarin stability (n=10).
3.3.4 miRNA association analysis

No association was found with miR-133a (miR-133a-3p, \( P=0.864 \), Figure 3.9A; miR-133a-5p, \( P=0.537 \), Figure 3.9B).

Interestingly, six miRNAs showed significant association (\( P<0.05 \)) with fold change >1.5 (Table 3.4). Three miRNAs were expressed at a significantly lower level in patients who achieved stable dose when compared to patients who did not achieve warfarin stability:

- SNORA31, \( P=0.0011 \), fold change =1.66;
- miR-5583-5p, \( P=0.0049 \), fold change =1.67;
- miR-548a-3p, \( P=0.0053 \), fold change =1.66.

Three miRNAs were expressed at a significantly higher level in the stable dose patients when compared to patients who did not achieve warfarin stability:

- SNORD114-12, \( P=0.0058 \), fold change =-1.61;
- miR-4643, \( P=0.0059 \), fold change =-1.57;
- miR-6750-3p, \( P=0.0090 \), fold change =-2.03.
Figure 3.9. Expression of (A) hsa-miR-133a-3p and (B) hsa-miR-133a-3p in serum from patients who have achieved warfarin stable dose (n=10), and from those who did not achieve warfarin stability (n=10). Adjusted $P$-values obtained using the MultiExperiment Viewer (MeV) v4.9 are shown in the upper right corner. Each dot represents an individual and solid lines represent the mean values.
Table 3.4. miRNA expression level in serum from warfarin-treated patients (n=20).

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>Stable dose (n=10)</th>
<th>Non-stable dose (n=10)</th>
<th>P-value</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNORA31</td>
<td>0.43±0.12</td>
<td>0.72±0.20</td>
<td>0.0011</td>
<td>1.68</td>
</tr>
<tr>
<td>hsa-miR-5583-5p</td>
<td>0.52±0.14</td>
<td>0.88±0.32</td>
<td>0.0049</td>
<td>1.67</td>
</tr>
<tr>
<td>hsa-miR-548a-3p</td>
<td>1.68±0.85</td>
<td>2.78±0.70</td>
<td>0.0053</td>
<td>1.66</td>
</tr>
<tr>
<td>SNORD114-12</td>
<td>0.88±0.29</td>
<td>0.54±0.18</td>
<td>0.0058</td>
<td>-1.61</td>
</tr>
<tr>
<td>hsa-miR-4643</td>
<td>0.77±0.27</td>
<td>0.49±0.09</td>
<td>0.0059</td>
<td>-1.57</td>
</tr>
<tr>
<td>hsa-miR-6750-3p</td>
<td>0.89±0.46</td>
<td>0.44±0.16</td>
<td>0.0090</td>
<td>-2.03</td>
</tr>
</tbody>
</table>

Data are expressed as average ± SD. Adjusted P-values were obtained using the MultiExperiment Viewer (MeV) v4.9. Fold changes were calculated as the ratio between the two conditions (non-stable dose versus stable dose).

Stable dose, patients who have achieved warfarin stable dose (n=10); non-stable dose, patients who did not achieve warfarin stability (n=10); SNORA, small nucleolar RNA, H/ACA box; SNORD, small nucleolar RNA, C/D box.
3.3.4.1 Bioinformatics analysis

Using the Affymetrix Transcriptome Analysis Console (TAC), v2.0, miR-548a-3p was found to target 1,042 genes. For the other five miRNAs which showed significant differential expression between patients who achieved warfarin stable dose and those who did not achieve warfarin stability, no target genes were found.

Interestingly, using the DAVID functional annotation tool, 6 target genes associated with miR-548a-3p were found to be involved in the coagulation pathway (Figures 3.10). The coagulation pathway with miR-548a-3p target genes is shown in Figure 3.11. miR-548a-3p regulates genes encoding for fibrinogen, coagulation factor V, and protein C. These genes are directly involved in the coagulation cascade (see section 1.4). In addition, miR-548a-3p regulates plasminogen, carboxypeptidase B2, and mannan-binding lectin serine peptidase 1 and 2 genes.

The protein encoded by plasminogen gene is the beta component of fibrinogen, a blood-borne glycoprotein composed of three pairs of non-identical polypeptide chains (Scott et al., 2004). Following vascular injury, fibrinogen is cleaved by thrombin to form fibrin, which is the most abundant component of blood clots (Doolittle, 1984).

Carboxypeptidase B2 gene encodes for enzymes that hydrolyze C-terminal peptide bonds (Tsai and Drayna, 1992). According to the substrate specificity, the enzyme is referred to as (i) carboxypeptidase A when it cleaves aliphatic residues, or (ii) carboxypeptidase B when it cleaves basic amino
residues. The protein encoded by this gene downregulates fibrinolysis after thrombin activation.

Mannan-binding lectin serine peptidase genes encodes several serine proteases that act mainly as components of the lectin pathway of complement activation. However, mannan-binding lectin-associated serine proteases (MASPs) exert several other functions (Yongqing et al., 2012). MASP1 has been reported to cleave fibrinogen, prothrombin and factor XIII, and to be capable of mediating fibrinogen polymerization (Jenny et al., 2015; Krarup et al., 2008), whereas MASP2 was shown to cleave prothrombin to activated factor X and hence promote clot formation (Krarup et al., 2007).

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>RT</th>
<th>Genes</th>
<th>Count</th>
<th>q</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEGG_PATHWAY</td>
<td>Cell cycle</td>
<td>RT</td>
<td>12</td>
<td>1.4</td>
<td>3.7E-2</td>
<td></td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>Pyruvate metabolism</td>
<td>RT</td>
<td>5</td>
<td>0.6</td>
<td>1.2E-1</td>
<td></td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>Splicesosome</td>
<td>RT</td>
<td>10</td>
<td>1.2</td>
<td>1.5E-1</td>
<td></td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>Nucleotide excision repair</td>
<td>RT</td>
<td>5</td>
<td>0.6</td>
<td>1.6E-1</td>
<td></td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>Purine metabolism</td>
<td>RT</td>
<td>11</td>
<td>1.3</td>
<td>2.0E-1</td>
<td></td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>Base excision repair</td>
<td>RT</td>
<td>4</td>
<td>0.5</td>
<td>2.3E-1</td>
<td></td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>Basal transcription factors</td>
<td>RT</td>
<td>4</td>
<td>0.5</td>
<td>2.3E-1</td>
<td></td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>Complement and coagulation cascades</td>
<td>RT</td>
<td>6</td>
<td>0.7</td>
<td>2.4E-1</td>
<td></td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>Androgen and estrogen metabolism</td>
<td>RT</td>
<td>4</td>
<td>0.5</td>
<td>2.6E-1</td>
<td></td>
</tr>
<tr>
<td>KEGG_PATHWAY</td>
<td>Biosynthesis of unsaturated fatty acids</td>
<td>RT</td>
<td>3</td>
<td>0.3</td>
<td>2.9E-1</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.10. Genes associated with miR-548a-3p.** Using the DAVID functional annotation tool, miR-548a-3p was predicted to be involved in the coagulation cascade.
Figure 3.11. Genes associated with miR-548a-3p are involved in the coagulation pathway. The red stars indicate genes associated with miR-548a-3p. CPB2, carboxypeptidase B2; FG, fibrinogen alpha chain, fibrinogen beta chain, fibrinogen gamma chain; F5, coagulation factor V; PROC, protein C; PLG, plasminogen; MASP1/2, mannan-binding lectin serine peptidase 1 and 2.
3.3.4.2 Identification of endogenous controls

Given that there is no universal endogenous control for miRNA qPCR experiments, we screened all the miRNAs (>6000) detected on the Affymetrix miRNA array for suitability. Two miRNAs, hsa-miR-92a-3p and hsa-miR-4487 (Figure 3.12A and 3.12B), were found to be unchanged between patients who achieved stable dose and patients who did not achieve warfarin stability ($P=0.996$ and $P=0.963$, respectively). As such, they were chosen as endogenous controls for downstream real-time qPCR experiments.

3.3.5 miRNA validation

To validate the findings from our pilot study, half of the remaining 980 prospectively recruited patients will be randomly selected as a validation cohort while the other 490 patients will serve as a replication cohort. miRNA-548a-3p as well as miR-133a-3p, will be assayed together with the two identified endogenous controls, miR-4487 and miR-92a-3p, and the technical spike-in control, cel-miR-39-3p. To ensure a thorough and even coverage of quantification range for each miRNA, standard curve optimisations were carried out.
Figure 3.12. Expression of (A) hsa-miR-92a-3p, and (B) hsa-miR-4487 in serum from patients who have achieved warfarin stable dose (n=10), and from those who did not achieve warfarin stability (n=10). Adjusted P-values obtained using the MultiExperiment Viewer (MeV) v4.9 are shown in the upper right corner. Each dot represents an individual and solid lines represent the mean values.
3.3.6  Real-time PCR standard curve optimisation

3.3.6.1  Phase 1

During phase 1 standard curve optimisation, prior to reverse transcription, the five synthetic miRNAs were serially diluted and five concentration points were chosen for downstream analysis, 1000 to 0.1 pM.

To assess PCR amplification efficiency, Ct values obtained from the standard curves were plotted against the logarithmic concentration of the input synthetic miRNA (Figure 3.13). The correlation coefficients of the slope and the amplification efficiency (indicated as %E) were calculated for all 5 miRNAs.

Although the correlation coefficients of all five standard curves ranged from 0.949 to 0.999, the amplification efficiency of all assays was very poor except for cel-miR-39 (cel-miR-39, %E = 104.85; miR-4487, %E = 61.38; miR-92a-3p, %E = 54.87; miR-133a-3p, %E = 59.19; miR-548a-3p, %E = 59.89). In addition, for miR-548a-3p, the last point of the standard curve (0.1 pM) was not detectable. Furthermore, at the 1 pM concentration point, miR-548a-3p showed a higher Ct value when compared to the other 4 miRNAs (cel-miR-39-3p, Ct = 20.78; miR-4487, Ct = 21.63; miR-92a-3p, Ct = 22.46; miR-133a-3p, Ct = 24.11; miR-548a-3p, Ct = 29.17).
Figure 3.13. Phase 1 standard curve optimisation. Serial dilutions (1000, 100, 10, 1, and 0.1 pM) of the five synthetic miRNAs were reverse-transcribed, pre-amplified and assayed using real-time PCR. The amplification curves showing delta Rn (ΔRn) versus Ct for synthetic cel-miR39-3p (A), miR4487 (B), miR-92a-3p (C), miR-133a-3p (D), and miR-548a-3p (E) are shown on the left while the fitted slope for Ct values versus respective the logarithmic concentration for each miRNA is shown on the right. Rn is the fluorescence of the reporter dye minus the fluorescence of a passive reference dye. The linear regression equation, the correlation coefficient of the line ($R^2$), and the efficiency of the assay (%$E = (10^{-1/slope} – 1) \times 100$) are shown in the upper right corner.
3.3.6.2 Phase 2

Given the poor amplification efficiency obtained in phase 1 optimisation, troubleshooting was carried out in an attempt to improve the amplification efficiency.

With consultation from technical support at Thermo Fisher Scientific, it was suggested that variability can be introduced at the reverse transcription and pre-amplification steps when different amounts of starting material are used. Advice was given to generate serially diluted standard curves after the pre-amplification step. As such, for each of the 5 concentrations which underwent reverse transcription and pre-amplification, 1:10 serial dilutions were carried out to generate a 6-point standard curve. With this procedure, five standard curves with different starting concentrations were generated for each miRNA.

Figure 3.14 details the $R^2$ and amplification efficiency of the standard curves generated for the two miRNAs, miR-133a-3p, and miR-548a-3p. A huge improvement in amplification efficiency was observed. The amplification efficiency for miR-133a-3p and miR-548a-3p ranged between 100.24-105.71% and 101.80-113.63%, respectively. In addition, the Ct of the two miRNAs at the same concentration were comparable. However, for miR-548a-3p, the last point of the standard curve (0.1 pM) was not detectable, suggesting that this particular miRNA assay could be slightly less sensitive than that for miR-133a-3p.
Figure 3.14. Phase 2 standard curve optimisation. A 100 µM stock for each miRNA of interest and four 1:10 serial dilutions from the stock (10 µM, 1 µM, 100 nM, 10nM) were reverse-transcribed and pre-amplified. For each pre-amplified product, 1:10 serial dilutions were carried out to generate a five-point standard curve. The amplification curves showing delta Rn (ΔRn) versus Ct for synthetic miR-133a-3p and miR-548a-3p are shown on the left while the fitted slopes for Ct values versus the logarithmic concentrations are shown on the right.

Six-point 1:10 serial dilutions ranged from 100 µM to 1 nM for mir-133a-3p (A) and miR-548a-3p (F); 10 µM to 100 pM for mir-133a-3p (B) and miR-548a-3p (G); 1 µM to 10 pM for mir-133a-3p (C) and miR-548a-3p (H); 100 nM to 1 pM for mir-133a-3p (D) and miR-548a-3p (I); 10 nM to 0.1 pM for mir-133a-3p (E) and miR-548a-3p (J).

The linear regression equation, the $R^2$, and the efficiency of the assay are shown in the upper right corner.
3.3.6.3 Phase 3

With the success from phase 2 optimisation, a concentration of 1 μM synthetic miRNA was chosen as the reverse transcription input amount. To capture the range between 5 and 35 Ct, an eight-point standard curve was generated using the pre-amplified product for each of the 5 miRNAs of interest via 1:10 serial dilution.

$R^2$ was >0.98 for all the five miRNAs and the amplification efficiency ranged between 91.50% to 98.96% (Figure 3.15), indicating high PCR efficiency.

Coefficient of variation (CV) was calculated for each miRNA assay for all the different concentrations in the 8-point standard curve. As shown in Table 3.5, intra-assay CV, expressed as the average across the eight calculated values, was less than 5% for all the assays, denoting accuracy and precision of the method.
Figure 3.15. Phase 3 standard curve optimisation. A 1 µM stock for each miRNA of interest was reverse-transcribed and pre-amplified. Seven serial dilutions (1:10) of the pre-amplified synthetic miRNAs were made, capturing a concentration range between 1 µM and 0.1 pM. The amplification curves showing delta Rn (ΔRn) versus Ct for synthetic cel-miR39-3p (A), miR4487 (B), miR-92a-3p (C), miR-133a-3p (D), and miR-548a-3p (E) are shown on the left while the fitted slope for Ct values versus the respective logarithmic concentration for each miRNA is shown on the right. The linear regression equation, the R², and the efficiency of the assay are shown in the upper right corner.
Table 3.5. Intra-assay coefficient of variation

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Average %CV</th>
<th>Min %CV</th>
<th>Max %CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>cel-miR-39-3p</td>
<td>1.47</td>
<td>0.35</td>
<td>2.68</td>
</tr>
<tr>
<td>miR-4487</td>
<td>1.30</td>
<td>0.39</td>
<td>3.05</td>
</tr>
<tr>
<td>miR-92a-3p</td>
<td>1.64</td>
<td>0.39</td>
<td>3.38</td>
</tr>
<tr>
<td>miR-133a-3p</td>
<td>1.86</td>
<td>0.27</td>
<td>4.85</td>
</tr>
<tr>
<td>miR-548a-3p</td>
<td>1.43</td>
<td>0.34</td>
<td>2.79</td>
</tr>
</tbody>
</table>

The coefficient of variation (CV) for each miRNA assay was calculated averaging the CV values across each standard curve. CV values are expressed as percentage. Minimum and maximum CV values for each assay are also reported.
3.4 DISCUSSION

In this chapter, we have conducted a pilot study to evaluate the effects of circulating miRNAs on warfarin dose variability. Genome-wide miRNA expression profiling was undertaken using serum samples from warfarinised patients and six miRNAs (SNORA31, miR-5583-5p, miR-548a-3p, SNORD114-12, miR-4643, and miR-6750-3p) exhibited significant differential expression when patients who achieved warfarin stable dose were compared to patients who did not achieve warfarin stability. Interestingly, among the six miRNAs, in silico analyses showed that several target genes of miR-548a-3p are involved in the coagulation pathway, suggesting that this circulating miRNA could play a role in warfarin dose variability, acting as a potential biomarker for warfarin treatment response. The expression level of miR-548a-3p was observed to be significantly higher in patients who did not achieve warfarin stable dose ($P=0.0053$, fold change =1.66) than patients who achieved stable dose. Currently, we do not know how miR-548a-3p regulates its target genes but it may be a combination of upregulation or downregulation processes. The functional role of miR-548a-3p in the coagulation cascade and its effect on warfarin response warrants further investigation.

Although genetic variants in miR-133a have been shown to affect warfarin dose variability (Ciccacci et al., 2015), we did not find any association between circulating serum miRNA-133a and warfarin response in our cohort of patients. Our small sample size may be underpowered to detect the small differences in miR133a expression levels between the stable dose and non-stable dose patients. Given that miRNA expression is tissue specific (Guo et
al., 2014; Liu et al., 2008b; Townley-Tilson et al., 2010), the reported association of miR-133a with VKORC1 mRNA by Pérez-Andreu and colleagues (2012) may be localized to liver tissue, and circulating miR-133a may not play a role in warfarin response. The true effect of circulating miR-133a on warfarin response will need to be validated in a larger cohort of patients.

Validation and replication of our preliminary findings will be carried out using Taqman miRNA real-time qPCR. Albeit circulating miRNAs offer many features such as stability, and evolutionarily conserved sequences that make them an attractive class of clinical biomarkers (Chen et al., 2008; Cortez and Calin, 2009; Lawrie et al., 2008; Mitchell et al., 2008), challenges associated with the detection of circulating miRNAs still need to be addressed. One of the challenges relates to the low yield of total RNA from plasma and serum, making it difficult to perform an accurate assessment of RNA quantity and quality. To correct the biases from variations in RNA input and reverse transcription efficiency, normalization with endogenous control genes is essential. The expression pattern of an ideal endogenous control should demonstrate relatively constant and highly abundant levels across all samples of the same cell type; however, no single control can serve as a universal endogenous control for all experimental conditions. From our Affymetrix miRNA array results, we identified two suitable endogenous controls specific to our study, miR-92a-3p and miR-4487.

To determine the absolute quantities of miRNA in our samples, standard curves were generated and optimised. Our optimisation results demonstrated that PCR amplification efficiency is highly affected by the
CHAPTER 3

performance of reverse transcription and pre-amplification. As such, serial dilution to generate standard curves should be carried out after the pre-amplification step to minimise poor efficiency. Carrying out serial dilutions prior to the pre-amplification step will lead to poor results.

Work is currently underway to validate and replicate the effects of circulating miR-548a-3p and miR-133a on warfarin response in our prospective cohort of warfarin-treated patients (n=980). miRNA from serum has already been extracted and quantified for all patients, with haemolysis evaluated. If these miRNAs are found to affect warfarin dose requirements, they will be incorporated into warfarin dosing algorithms to determine their effect on the accuracy of warfarin dosing. Functional work to elucidate the miRNA effects on target genes will also be carried out.
CHAPTER 4

Warfarin Extreme Phenotypes: A Genome-Wide Association Study
CHAPTER 4

CONTENTS

4.1 INTRODUCTION ................................................................. 143

4.2 MATERIALS AND METHODS ........................................... 149
  4.2.1 Subjects ........................................................................... 149
  4.2.2 Genome-wide genotyping ............................................. 150
  4.2.3 Data quality control ...................................................... 152
    4.2.3.1 Sample quality control ........................................ 152
    4.2.3.2 SNP quality control ............................................... 154
  4.2.4 Phasing and Imputation ................................................ 154
  4.2.5 Statistical analysis ....................................................... 155

4.3 RESULTS ................................................................. 156
  4.3.1 Patient characteristics .................................................. 156
  4.3.2 Univariate association analysis of clinical variables in patients
    sensitive to warfarin versus patients resistant to warfarin ....... 156
  4.3.3 Genome-wide association analysis ................................. 160
    4.3.3.1 Warfarin sensitive versus warfarin resistant patients ... 160
    4.3.3.2 Warfarin extreme cases versus healthy controls ........... 171
    4.3.3.3 Warfarin sensitivity versus healthy controls ............... 171
    4.3.3.4 Warfarin resistance versus healthy controls ............... 179

4.4 DISCUSSION .................................................................. 185
CHAPTER 4

4.1 INTRODUCTION

Warfarin pharmacogenetic dosing algorithms incorporating the genetic variants VKORC1 -1639G>A (rs9923231), CYP2C9*2 (rs1799853), and CYP2C9*3 (rs1057910), together with clinical variables such as age and BMI have been shown to improve warfarin dosing accuracy (Anderson et al., 2007; Gage et al., 2008; Klein et al., 2009; Limdi et al., 2010). However, the pharmacogenetic dosing approach under predicts patients who require extremely high doses of warfarin and over predicts patients who need very low doses of warfarin (Gage et al., 2008; Klein et al., 2009). This suggests that common variants in VKORC1 and CYP2C9 may not explain the rare cases of very high and low doses of warfarin required in patients with warfarin resistance and sensitivity, respectively.

Warfarin resistance has been described as the inability to prolong the prothrombin time or raise the INR into the therapeutic range when the drug is administered at a dose near or equivalent to the normally recommended doses (Lefrere et al., 1987). The prevalence of warfarin resistance varies by ethnicity and is difficult to determine. However, a higher warfarin requirement does not itself establish the diagnosis of warfarin resistance. The difficulty lies largely in accounting for dietary factors and in defining normal metabolic variations among individuals. Transient causes of acquired resistance to warfarin include poor patient compliance (the most common cause), excessive dietary vitamin K intake (O’Reilly and Rytand, 1980), drug interactions (MacLeod and Sellers, 1976), and decreased absorption of warfarin.

In mechanistic terms, warfarin resistance can be classified into pharmacokinetic resistance and pharmacodynamic resistance.
Pharmacokinetic resistance results from diminished absorption or increased elimination of the drug. Causes of diminished absorption include emesis, diarrhoea, and malabsorption syndrome. Enhanced warfarin metabolism arises predominantly through the induction of CYP2C9 and to a lesser extent, CYP1A1, CYP1A2 and CYP3A4. Duplication or multiplication of cytochrome P450 enzyme genes has been described as contributing to a phenotype of ultrarapid metabolism, as it has already been reported for CYP2D6 with debrisoquine, tramadol and codeine, and CYP2A6 with nicotine (Daly and Aithal, 2003; Gasche et al., 2004; Johansson et al., 1993; Kirchheiner et al., 2008; Takahashi and Echizen, 2001). However, no duplication has yet been reported for CYP2C9, CYP1A1, CYP1A2 and CYP3A4.

Pharmacodynamic warfarin resistance occurs when an elevated circulating warfarin concentration beyond that of the upper limit of the therapeutic range is achieved through high daily doses of warfarin, while the INR remains consistently sub-therapeutic. The mechanism of pharmacodynamic warfarin resistance in man has not been clearly delineated; postulated mechanisms include increased affinity of VKORC1 for vitamin K (Cain et al., 1998; O’Reilly et al., 1968), decreased VKORC1 sensitivity to warfarin (Cain et al., 1998), prolongation of normal clotting factor activity and production of non-vitamin K-dependent clotting factors (O’Reilly, 1970).

Several rare missense mutations in the VKORC1 gene are responsible for resistance to warfarin, as summarised in Table 4.1. However, these studies were either isolated case reports (Ainle et al., 2008; D’Ambrosio et al., 2007; Harrington et al., 2005; Loebstein et al., 2007; Rost et al., 2004a) or carried out in a small number of patients (Harrington et al., 2008; Watzka et
al., 2011). A missense mutation in VKORC1 gene associated with warfarin resistance, Asp36Tyr (rs61742245), is confined to North-Eastern African and Middle-Eastern populations. This mutation was found to be present in 15% of the Ethiopians with a significant influence on warfarin dose requirement (>70 mg/week), accounting for 5.5% warfarin dose variability (Loebstein et al., 2007; Shahin et al., 2013).

By contrast, warfarin sensitivity is a condition in which individuals have a low tolerance for the drug. The prevalence of warfarin sensitivity is unknown. Nevertheless, it appears to be more common in people who are older (see section 1.5), those with lower body weight, and individuals of Asian ancestry. Genetic associations of VKORC1 -1639G>A, CYP2C9*2, CYP2C9*3, and CYP4F2 with warfarin sensitivity has been reported in numerous studies and have been previously discussed in sections 1.7.1, 1.7.2 and 1.7.3.

Several additional mutations in the CYP2C9 gene have been reported in patients sensitive to warfarin (see Table 4.2). However, given the low frequency of these polymorphisms, they have minimal impact on warfarin dose requirements in some but not all the populations tested. Furthermore, polymorphisms in ABCB1 (rs1045642) (Wadelius et al., 2004), factor II (Thr165Met, rs5896) (D’Ambrosio et al., 2004; Shikata et al., 2004), and factor VII (Arg353Gln) (Aquilante et al., 2006; D’Ambrosio et al., 2004; Mlynarsky et al., 2012; Shikata et al., 2004) genes can also contribute to increased warfarin sensitivity.

To identify additional genetic variants contributing to warfarin resistance and sensitivity, in this experimental chapter we performed a
genome-wide association study in patients requiring very high and low doses of warfarin.
Table 4.1. Previously reported *VKORC1* mutations in warfarin resistant patients.

<table>
<thead>
<tr>
<th>Mutationa</th>
<th>Nucleotideb</th>
<th>dbSNP rs number</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>p.Ala26Pro</td>
<td>c.76G&gt;C</td>
<td>-</td>
<td>(Bodin <em>et al.</em>, 2005a)</td>
</tr>
<tr>
<td>p.Val29Leu</td>
<td>c.85G&gt;T</td>
<td>rs104894539</td>
<td>(Rost <em>et al.</em>, 2004a)</td>
</tr>
<tr>
<td>p.Ala41Ser</td>
<td>c.121G&gt;T</td>
<td>-</td>
<td>(Bodin <em>et al.</em>, 2005b; Harrington <em>et al.</em>, 2008)</td>
</tr>
<tr>
<td>p.Val45Ala</td>
<td>c.134T&gt;C</td>
<td>rs104894540</td>
<td>(Rost <em>et al.</em>, 2004a)</td>
</tr>
<tr>
<td>p.Arg58Gly</td>
<td>c.172G&gt;C</td>
<td>rs104894541</td>
<td>(Rost <em>et al.</em>, 2004a)</td>
</tr>
<tr>
<td>p.Val66Met</td>
<td>c.196G&gt;A</td>
<td>rs72547529</td>
<td>(Watzka <em>et al.</em>, 2011)</td>
</tr>
<tr>
<td>p.Leu128Arg</td>
<td>c.383T&gt;G</td>
<td>rs104894542</td>
<td>(Rost <em>et al.</em>, 2004a)</td>
</tr>
<tr>
<td>p.Arg151Gln</td>
<td>c.452G&gt;A</td>
<td>-</td>
<td>(Rost <em>et al.</em>, 2004a)</td>
</tr>
</tbody>
</table>

*a* Amino acid numbering is according to reference sequence NP_076869.

*b* Nucleotide numbering is according to reference sequence NM_024006.
Table 4.2. Previously reported CYP2C9 mutations in warfarin sensitive patients.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Nucleotide</th>
<th>dbSNP rs number</th>
<th>Allele</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>273frameshift</td>
<td>c.818delA</td>
<td>rs9332131</td>
<td>*6</td>
<td>(Sagrieya et al., 2010)</td>
</tr>
<tr>
<td>p.Arg150His</td>
<td>c.449G&gt;A</td>
<td>rs7900194</td>
<td>*8</td>
<td>(Liu et al., 2012)</td>
</tr>
<tr>
<td>p.Arg335Trp</td>
<td>c.1003C&gt;T</td>
<td>rs28371685</td>
<td>*11</td>
<td>(Sagrieya et al., 2010)</td>
</tr>
<tr>
<td>p.Pro489Ser</td>
<td>c.1465C&gt;T</td>
<td>rs9332239</td>
<td>*12</td>
<td>(O’Brien et al., 2013; Sagrieya et al., 2010)</td>
</tr>
<tr>
<td>p.Leu90Pro</td>
<td>c.269T&gt;C</td>
<td>rs72558187</td>
<td>*13</td>
<td>(Di Fusco et al., 2013)</td>
</tr>
<tr>
<td>p.Arg125His</td>
<td>c.374G&gt;A</td>
<td>rs72558189</td>
<td>*14</td>
<td>(Ciccacci et al., 2011; DeLozier et al., 2005; Kim et al., 2009a; Zhao et al., 2004)</td>
</tr>
<tr>
<td>p.Ser162Leu</td>
<td>c.485C&gt;A</td>
<td>rs72558190</td>
<td>*15</td>
<td>(Zhao et al., 2004)</td>
</tr>
<tr>
<td>p.Thr299Ala</td>
<td>c.895A&gt;G</td>
<td>rs72558192</td>
<td>*16</td>
<td>(Zhao et al., 2004)</td>
</tr>
<tr>
<td>p.Asp397Ala</td>
<td>c.1190A&gt;C</td>
<td>rs72558193</td>
<td>*18</td>
<td>(Sagrieya et al., 2010)</td>
</tr>
<tr>
<td>p.Arg125Leu</td>
<td>c.374G&gt;T</td>
<td>-</td>
<td>*35</td>
<td>(Ciccacci et al., 2011; Lee et al., 2014)</td>
</tr>
<tr>
<td>p.Pro337Thr</td>
<td>c.1009C&gt;A</td>
<td>-</td>
<td>*58</td>
<td>(Luo et al., 2014)</td>
</tr>
</tbody>
</table>

a Amino acid numbering is according to reference sequence NP_000762.

b Nucleotide numbering is according to reference sequence NM_000771.
4.2 MATERIALS AND METHODS

4.2.1 Subjects

Patients showing resistance or sensitivity to warfarin were recruited both prospectively and retrospectively from the Royal Liverpool University Hospital and University Hospital Aintree. The study was approved by the Birmingham South research ethics committee and written informed consent was obtained from all patients. Among the 116 recruited, 60 were resistant to warfarin, 56 were sensitive to warfarin.

Warfarin resistance was defined as the requirement of ≥10 mg warfarin per day on three successive clinic visits (prospectively recruited patients) or for at least three weeks (retrospectively recruited patients), in the absence of enzyme inducers (carbamazepine, phenytoin, phenobarbitone, rifampicin) (retrospectively recruited patients).

Warfarin sensitivity was defined as the requirement of ≤1.5 mg warfarin per day on three successive clinic visits (prospectively recruited patients) or for at least three weeks (retrospectively recruited patients) in the absence of enzyme inhibitors (amiodarone, omeprazole, simvastatin, fluoxetine).

Genotypic data from 2,501 healthy blood donors from the National Blood Service (NBS) were used as healthy controls for GWAS. Access to raw data from the control data set for the study was granted by The Wellcome Trust Case-Control Consortium (to Dr Eunice Zhang).
4.2.2 Genome-wide genotyping

DNA samples from the 116 patients were shipped to Edinburgh-Genomics (The Roslin Institute, University of Edinburgh) for genome-wide genotyping on the Illumina Human OmniExpressExome-8v1 BeadChip (n=4), Illumina Human OmniExpressExome-8v1.1 BeadChip (n=16), and Illumina HumanExome-12v1 BeadChip (n=96) using the Infinium HD Super assay (Illumina, Inc.).

Figure 4.1 summarises the principles and chemistry of genome-wide genotyping on the Illumina platform. In this system, whole-genome amplification is used to increase the amount of DNA up to 1000-fold. The DNA is fragmented and captured on a BeadChip by hybridisation to immobilised SNP-specific primers, stopping one base before the locus of interest. Marker specificity is conferred by enzymatic single-base extension where a fluorescently-labelled nucleotide is incorporated. Subsequent dual-colour fluorescent staining allows the labelled nucleotide to be detected by Illumina’s iScan imaging system, which identifies both colour and signal intensity.

951,117 SNPs were mapped to NCBI build 36 using Illumina Human OmniExpressExome-8v1.0_B.bpm manifest file (4 samples), 958,178 SNPs were mapped using Human OmniExpressExome-8v1.1_C.bpm manifest file (16 samples), while 242,901 SNPs were mapped using HumanExome-12v1-1_A.bpm manifest file (96 samples). The number of overlapping markers was 235,403, of which 135 were insertion/deletions, and 235,268 were bi-allelic SNPs. Bead intensity data were processed and normalized for each sample in GenomeStudio Data Analysis Software v1.0 (Illumina, Inc.).
Figure 4.1. Outline of the Infinium HD assay chemistry for whole-genome genotyping. Five modular steps are involved. (i) DNA Amplification. After denaturation and neutralisation, the DNA is isothermally amplified whereby whole-genome amplification uniformly increases the amount of DNA by several thousand-fold without introducing large amount of amplification bias. (ii) Fragmentation and precipitation. The amplified product is fragmented by a controlled enzymatic endpoint fragmentation process to avoid over-fragmenting the sample. Following isopropanol precipitation, the fragmented DNA is resuspended in hybridization buffer. (iii) Hybridisation to BeadChip. The BeadChip is prepared for hybridisation in a capillary flow-through chamber. During hybridisation, the amplified and fragmented DNA samples selectively anneal to locus-specific 50-mer probes, stopping one base before the interrogated marker. Unhybridised and non-specifically hybridised DNA is washed away. (iv) Single base extension and staining. If there is a perfect match, extension occurs. Extension of the primer incorporates a biotin nucleotide or a dinitrophenyl-labelled nucleotide. C and G nucleotide are biotin labelled (green); A and T nucleotides are dinitrophenyl-labelled (red). Signal amplification of the incorporated label further improves the overall signal-to-noise ratio of the assay. (v) Array scanning and genotype calling. The Illumina BeadArray Reader uses a laser to excite the fluorophore of the single-base extension product on the beads of the BeadChip sections, recording high-resolution images of light emitted from the fluorophores. Data were analysed using Illumina’s GenomeStudio Data Analysis Software v1.0 (Illumina, Inc.).
4.2.3 Data quality control

Quality control (QC) procedures of the genome-wide genotype data was carried out at both the sample and SNP level using the freely available software PLINK v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al., 2007).

4.2.3.1 Sample quality control

The reported gender of each patient was checked against that predicted by the genetic data using X chromosome homozygosity estimates. Three samples were found to be gender discrepant and were removed.

Individuals with genotyping call rate ≤ 95% were excluded (n=3). Given that excessive proportion of heterozygote genotypes may be indicative of DNA sample contamination while reduced proportion of heterozygote genotypes may suggest inbreeding, the heterozygosity rate for each individual was calculated. One sample was removed due to low heterozygosity (<0.16).

In order to identify duplicated and related individuals, the identity-by-descent (IBD) metric was calculated for each pairwise sample combination. Two individuals showed cryptic relationship as first degree relative (IBD >0.5, \(\text{pi-Hat} >0.2\)). Between the two samples, the one with the higher genotype call rate was retained, and the other was excluded.

The ethnic origin of the patients was confirmed by principal-component analysis (PCA). HapMap release 23 for Caucasian (CEU), Han Chinese (CHB), Japanese (JPT), and Yoruban (YRI) populations were used as reference populations. Samples were identified as ethnic outliers on the
basis of their projection onto the four principal components of genetic variations (Figure 4.2). All subjects in our cohort clustered with the HapMap CEU samples except for two subjects showing evidence of Yoruban ancestry (see Figure 4.2) and were excluded. In total 10 samples failed QC. Of the remaining 106 patients, 51 were resistant to warfarin and 55 were sensitive to warfarin.

Figure 4.2. PCA plot of 116 Warfarin Extreme Phenotype study samples. CEU, HapMap Caucasians (n=165); CHB, HapMap Han Chinese (n=84); JPT, Hapmap Japanese (n=86); YRI, HapMap Yoruban (n=167); WEP, Warfarin Extreme Phenotype study samples (n=116). The red arrows indicate ethnic outliers.
4.2.3.2 SNP quality control

SNPs which deviate from Hardy–Weinberg Equilibrium (HWE) with $P<1\times10^{-5}$ (n=391) were removed. SNPs with minor allele frequency less than 1% (n=14,518) and with call rate ≤95% (n=199,672) were excluded from downstream analysis. Of the 235,403 SNPs, 31,596 SNPs passed QC.

4.2.4 Phasing and Imputation

After GWAS QC, SNP alignment was undertaken. The reference genotype data (CEU) from the 1000 Genomes project (Phase I integrated variant set release v3, NCBI build 37 (hg19)) were used (1000 Genomes Project Consortium et al., 2010; Price et al., 2006). Mismatched SNP alleles and variants not present in the 1000 Genomes project were removed. Non-autosomal SNPs (n=274) and those that gave A>T or G>C or vice versa (n=3,013) were excluded from the dataset, and 28,309 SNPs remained.

Haplotype phasing was carried out using the software SHAPEIT v2.r644 (Delaneau et al., 2013). Imputation of additional SNPs on chromosomes 1 to 22 throughout the whole genome was performed using the program IMPUTE v2.3.0 (Howie et al., 2012, 2009; Marchini et al., 2007).

A total of 859,028 SNPs were imputed. Post-imputation SNP QC was carried using QCTool and PLINK. Imputed variants with low imputation quality ($r^2<0.8$), genotyping success rate ≤95%, MAF <1% and HWE $P<1\times10^{-5}$ were excluded and 558,416 SNPs remain for analysis.
### 4.2.5 Statistical analysis

Statistical analyses were conducted in SPSS v22.0 or in Plink v1.07. To test the differences in clinical and demographic features among patients, the Student’s t-test and analysis of variance (ANOVA) were undertaken for continuous variables. For categorical variables, the Pearson's chi-square test was used. Clinical variables with a $P<0.05$ were included as covariates in the regression analysis. To avoid collinearity, correlation between each pair of clinical factors was assessed, with only one from each highly correlated pair included as covariate.

Logistic regression was employed to test for SNP associations with binary outcomes, namely warfarin resistant versus warfarin sensitive, warfarin resistant versus healthy controls, warfarin sensitive versus healthy controls.

Any SNP showing a $P$-value $<5\times10^{-8}$ was regarded as genome-wide significant. A threshold of $P$-value $<1\times10^{-5}$ indicated suggestive genome-wide significance. Conditional analysis of SNPs was undertaken using Plink v1.07.

Manhattan plots were generated using the genetics package of R, version 3.1.1 (http://cran.rproject.org/web/packages/genetics/index.html). The open-source web-based plotting tool, LocusZoom (Pruim et al., 2010) was used for regional visualisation of genome-wide results.

Analysis of LD (Gabriel et al., 2002) was performed using the Haploview software, version 4.2 (Barrett et al., 2005).
4.3 RESULTS

4.3.1 Patient characteristics

The clinical, anamnestic and demographic features of 106 warfarin-treated patients (51 resistant to warfarin and 55 sensitive to warfarin) are shown in Table 4.3. The mean age was 66 years (range 27-100 years), and 53.8% were men. The main indication for warfarin therapy was atrial fibrillation (54.9%).

Information on comorbidities and concomitant medications were gathered from all patients. All adverse events during the study were also recorded, and 14.9% of patients experienced an adverse event, demonstrating the fragility of the considered phenotypes. Patients sensitive to warfarin showed higher incidences of adverse events (minor bleedings) than patients resistant to warfarin (27.0% versus 2.7%). Indeed, patients requiring very low doses of warfarin are at risk of overtreatment, which can lead to bleeding events (Klein et al., 2009; Moyer et al., 2009).

For the 2,501 healthy controls from NBS, only gender information was available and 49.5% were men.

4.3.2 Univariate association analysis of clinical variables in patients sensitive to warfarin versus patients resistant to warfarin

Table 4.3 summarises univariate analyses results of clinical variables when warfarin sensitive patients were compared to warfarin resistant patients.
In accordance with some reports which demonstrated that males require higher warfarin doses (Absher et al., 2002; Anderson et al., 2007; Cini et al., 2012), our study showed that a significantly higher proportion of male patients were resistant to warfarin when compared to patients sensitive to warfarin ($P=0.003$). In patients resistant to warfarin, the body surface area was significantly higher when compared to patients sensitive to warfarin ($P=0.002\times10^{-3}$). Indeed, this finding is consistent with several studies (Klein et al., 2009; Wadelius et al., 2009; Zambon et al., 2011) where increased BSA was associated with higher warfarin requirements. The percentage of patients indicated for warfarin therapy due to atrial fibrillation was higher in the group of patients sensitive to warfarin compared to the patients resistant to warfarin (62.9% versus 40.0%; $P=0.007$).

To avoid collinearity, correlation between clinical variables that showed significant univariate associations was assessed. Gender was associated with BSA ($P=1.793\times10^{-8}$; Figure 4.3), where BSA was significantly higher in males than females. No correlation was found between gender and indication for atrial fibrillation ($P=0.129$), or between BSA and indication for atrial fibrillation ($P=0.497$). Therefore, BSA, which showed a stronger correlation compared to gender, was included as a covariate together with the indication for atrial fibrillation in our logistic regression analysis.

No significant differences in comorbidities and concomitant medications were found between the sensitive and resistant patients.
Table 4.3. Clinical, amnestic and demographic features of the warfarin-treated patients (n=106).

<table>
<thead>
<tr>
<th></th>
<th>All (106)</th>
<th>Sensitive (55)</th>
<th>Resistant (51)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>66.28±14.68</td>
<td>67.27±14.00</td>
<td>65.22±15.44</td>
<td>0.474</td>
</tr>
<tr>
<td>Gender (male)</td>
<td></td>
<td></td>
<td></td>
<td>0.003</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>29.55±6.43</td>
<td>28.61±6.22</td>
<td>30.53±6.56</td>
<td>0.133</td>
</tr>
<tr>
<td>BSA (m²)</td>
<td>1.97±0.30</td>
<td>1.84±0.25</td>
<td>2.11±0.28</td>
<td>0.002x10⁻³</td>
</tr>
<tr>
<td>Indication for warfarin therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>54.9%</td>
<td>69.2%</td>
<td>40.0%</td>
<td>0.007</td>
</tr>
<tr>
<td>Deep venous thrombosis</td>
<td>18.6%</td>
<td>11.5%</td>
<td>26.0%</td>
<td>0.051</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>14.7%</td>
<td>9.6%</td>
<td>20.0%</td>
<td>0.121</td>
</tr>
<tr>
<td>CVA/TIA</td>
<td>7.8%</td>
<td>7.7%</td>
<td>8.0%</td>
<td>0.912</td>
</tr>
<tr>
<td>Cardiac valve replacement</td>
<td>3.9%</td>
<td>1.9%</td>
<td>6.0%</td>
<td>0.273</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>72.4%</td>
<td>72.7%</td>
<td>72.0%</td>
<td>0.934</td>
</tr>
<tr>
<td>Respiratory disease</td>
<td>40.6%</td>
<td>47.3%</td>
<td>33.3%</td>
<td>0.114</td>
</tr>
<tr>
<td>Hepatic disease</td>
<td>6.7%</td>
<td>3.6%</td>
<td>10.0%</td>
<td>0.192</td>
</tr>
<tr>
<td>Gastrointestinal disorder</td>
<td>18.9%</td>
<td>25.5%</td>
<td>11.8%</td>
<td>0.072</td>
</tr>
<tr>
<td>Renal disease</td>
<td>14.3%</td>
<td>16.4%</td>
<td>12.0%</td>
<td>0.523</td>
</tr>
<tr>
<td>Urological condition</td>
<td>20.0%</td>
<td>27.3%</td>
<td>12.0%</td>
<td>0.051</td>
</tr>
<tr>
<td>Neurological disorder</td>
<td>11.3%</td>
<td>16.4%</td>
<td>5.9%</td>
<td>0.089</td>
</tr>
<tr>
<td>Concomitant drug consumption</td>
<td>90.6%</td>
<td>96.4%</td>
<td>84.3%</td>
<td>0.052</td>
</tr>
</tbody>
</table>

Data are expressed as average ± SD (continuous variables) or in percentage (categorical variables). CVA, cardiovascular accident; TIA, transient ischemic attack.
Figure 4.3. Relationship between gender and BSA. Box and whisker plots showing the distribution of BSA among males and females. Boxes represent the 25th-75th percentiles (interquartile range), whiskers represent 5th-95th percentiles, solid lines represent the median values in each group and dots represent outliers. The $P$-value, calculated by ANOVA is shown in the upper right corner.
4.3.3 Genome-wide association analysis

4.3.3.1 Warfarin sensitive versus warfarin resistant patients

When warfarin sensitive patients were compared to patients with warfarin resistance, taking BSA and warfarin indication as clinical covariates, no genome-wide association was observed. However, two signals suggestive of genome-wide significance were identified. One tower clustered on chromosome 10 while the other on chromosome 16 (Figure 4.4). The lists of top SNPs which attained suggestive genome-wide P-values are shown in Tables 4.4 and 4.5.

On chromosome 10, the top signal was an imputed SNP, rs4918797, \( (P=2.97\times10^{-6}; \text{Figure 4.5A}) \). rs4918797 is a C/T single-nucleotide intergenic variant located approximately 1100bp from the 3’ end of the \( CYP2C9 \) gene. The minor T allele at rs4918797 was present only in patients sensitive to warfarin and none of the resistant patients was a TT-carrier (Table 4.6). Of note, \( CYP2C9^*2 \) and \( ^*3 \) showed \( P \)-values below the suggestive threshold of \( 1\times10^{-5} \) \( (P=0.074 \text{ and } P=6.15\times10^{-4}, \text{respectively}) \). Interestingly, rs4918797 displayed linkage disequilibrium with \( CYP2C9^*2 \) \( (D'=1.0, r^2=0.345) \) and \( CYP2C9^*3 \) \( (D'=1.0, r^2=0.437) \) (Figure 4.6), with \( D' \) values of 1, suggesting these three SNPs have not been separated by recombination over generations.

On chromosome 16, the strongest signal was rs71383768 \( (P=3.93\times10^{-6}; \text{Figure 4.5B}) \), an imputed SNP. rs71383768 is an intergenic variant located between the \( LOC101928762 \) gene and \( ZNF668 \) which encodes for the zinc finger protein 668. The \( LOC101928762 \) gene is affiliated
with the non-coding RNA (ncRNA) class. The function of ZNF668 gene is not clear (Hu et al., 2011).

Interestingly, no association was found with the VKORC1 3’UTR 3730 G>A SNP, rs7294 ($P=0.066$), suggesting that common variants in VKORC1 do not explain the extreme doses required by patients sensitive and resistant to warfarin (Gage et al., 2008; Klein et al., 2009).

The minor C allele at rs71383768 was observed predominantly in patients sensitive to warfarin, with a CC genotype frequency of 63.5% while that in patients resistant to warfarin was only 2% (Table 4.6).

To check for SNP independence, conditional analyses were performed using rs4918797 and rs71383768. After conditioning for rs4918797 on chromosome 10, the tower of SNPs on chromosome 10 disappeared (Figure 4.7A), and no SNPs reached suggestive genome-wide significance. The top signal on chromosome 16 changed from rs71383768 to rs11150601 ($P=1.73\times10^{-4}$, Figure 4.7B), an intronic SNP located in the SETD1A (SET domain containing 1A) gene. Likewise, after conditioning for rs71383768 on chromosome 16 (Figure 4.8A), rs4918797 did not remain as the top SNP on chromosome 10 ($P=4.30\times10^{-04}$, Figure 4.8B). The new top SNP on chromosome 10 was rs2860905 ($P=3.32\times10^{-04}$, Figure 4.8B), an A/G single-nucleotide intronic variant located approximately 1100bp from the 3’ end of the CYP2C9 gene. Furthermore, after conditioning for both rs4918797 and rs71383768, no SNPs reached suggestive genome-wide significance, suggesting that rs4918797 and rs71383768 are not independent of each other (Figure 4.9).
Figure 4.4. Genome-wide Manhattan plot of warfarin sensitive patients versus warfarin resistant patients, with BSA and warfarin indication included as clinical covariates. Individual –log_{10} P-values are plotted against their genomic position by chromosome. The top red line at 5x10^{-8} marks the genome-wide significant threshold, while the blue line at 10^{-5} marks the threshold for promising SNPs. The green dots depict genotyped SNPs, and the red dots depict imputed SNPs.
Figure 4.5. Regional association plots of chromosomes 10 (A) and 16 (B). Sensitive (n=55) versus resistant (n=51) patients were compared. The target SNPs are depicted with a purple diamond. Each coloured circle represents a SNP $P$-value, with the colour scale reflecting the extent of LD ($r^2$). SNPs with missing LD information are shown in grey. Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on build 37 (NCBI) of the human genome. The relative positions of gene mapping to the region of association are also shown.
Table 4.4. Top 10 SNPs on chromosome 10 showing suggestive $P$-values in the outcome warfarin sensitivity versus warfarin resistance.

<table>
<thead>
<tr>
<th>Variant ID</th>
<th>Position</th>
<th>Locus</th>
<th>Localisation</th>
<th>Alleles</th>
<th>MAF Sensitive</th>
<th>MAF Resistant</th>
<th>Reported MAF*</th>
<th>$P$</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4918797</td>
<td>96750251</td>
<td>CYP2C9, CYP2C59P</td>
<td>Intergenic</td>
<td>C/T</td>
<td>0.582</td>
<td>0.080</td>
<td>0.199</td>
<td>2.97e-06</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs4918798</td>
<td>96750543</td>
<td>CYP2C9, CYP2C59P</td>
<td>Intergenic</td>
<td>C/A</td>
<td>0.582</td>
<td>0.080</td>
<td>0.199</td>
<td>2.97e-06</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs4086116</td>
<td>96707202</td>
<td>CYP2C9</td>
<td>Intrinsic</td>
<td>C/T</td>
<td>0.582</td>
<td>0.088</td>
<td>0.198</td>
<td>3.27e-06</td>
<td>Genotyped</td>
</tr>
<tr>
<td>rs1934963</td>
<td>96734676</td>
<td>CYP2C9</td>
<td>Intrinsic</td>
<td>T/C</td>
<td>0.582</td>
<td>0.088</td>
<td>0.200</td>
<td>3.27e-06</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs4917639</td>
<td>96725535</td>
<td>CYP2C9</td>
<td>Intrinsic</td>
<td>T/C</td>
<td>0.582</td>
<td>0.088</td>
<td>0.198</td>
<td>3.27e-06</td>
<td>Genotyped</td>
</tr>
<tr>
<td>rs6188604</td>
<td>96749936</td>
<td>CYP2C9</td>
<td>Intrinsic</td>
<td>T/G</td>
<td>0.582</td>
<td>0.088</td>
<td>0.200</td>
<td>3.27e-06</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs9332172</td>
<td>96731788</td>
<td>CYP2C9</td>
<td>Intrinsic</td>
<td>A/G</td>
<td>0.582</td>
<td>0.088</td>
<td>0.200</td>
<td>3.27e-06</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs9332220</td>
<td>96743943</td>
<td>CYP2C9</td>
<td>Intrinsic</td>
<td>G/A</td>
<td>0.582</td>
<td>0.088</td>
<td>0.198</td>
<td>3.27e-06</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs9332238</td>
<td>96748492</td>
<td>CYP2C9</td>
<td>Intrinsic</td>
<td>G/A</td>
<td>0.582</td>
<td>0.088</td>
<td>0.197</td>
<td>3.27e-06</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs74494115</td>
<td>96719845</td>
<td>CYP2C9</td>
<td>Intrinsic</td>
<td>G/T</td>
<td>0.583</td>
<td>0.080</td>
<td>0.198</td>
<td>3.39e-06</td>
<td>Imputed</td>
</tr>
</tbody>
</table>

*Physical positions are based on build 37 (NCBI) of the human genome.
*1000 Genomes, Europeans.
CYP2C59P, pseudogenes.
Table 4.5. Top 10 SNPs on chromosome 16 showing suggestive $P$-values in the outcome warfarin sensitivity versus warfarin resistance.

<table>
<thead>
<tr>
<th>Variant ID</th>
<th>Position*</th>
<th>Locus</th>
<th>Localisation</th>
<th>Alleles</th>
<th>MAF Sensitive</th>
<th>MAF Resistant</th>
<th>Reported MAF*</th>
<th>$P$</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs71383768</td>
<td>31063979</td>
<td>LOC101928762,ZNF668</td>
<td>Intergenic</td>
<td>T/C</td>
<td>0.741</td>
<td>0.200</td>
<td>-</td>
<td>3.93x10^{-6}</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs11076</td>
<td>30995528</td>
<td>HSD3B7</td>
<td>Upstream</td>
<td>A/G</td>
<td>0.873</td>
<td>0.451</td>
<td>0.634</td>
<td>6.66x10^{-6}</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs8060857</td>
<td>31002720</td>
<td>STX1B</td>
<td>Upstream</td>
<td>G/A</td>
<td>0.889</td>
<td>0.441</td>
<td>0.635</td>
<td>7.06x10^{-6}</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs8062719</td>
<td>31002664</td>
<td>STX1B</td>
<td>Upstream</td>
<td>A/G</td>
<td>0.889</td>
<td>0.440</td>
<td>0.646</td>
<td>7.39x10^{-6}</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs9938550</td>
<td>30999142</td>
<td>HSD3B7</td>
<td>Synonymous</td>
<td>A/G</td>
<td>0.873</td>
<td>0.441</td>
<td>0.636</td>
<td>8.03x10^{-6}</td>
<td>Genotyped</td>
</tr>
<tr>
<td>rs11150601</td>
<td>30977799</td>
<td>SETD1A</td>
<td>Intronic</td>
<td>G/A</td>
<td>0.873</td>
<td>0.441</td>
<td>0.635</td>
<td>8.03x10^{-6}</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs12716978</td>
<td>30981544</td>
<td>SETD1A</td>
<td>Intronic</td>
<td>C/G</td>
<td>0.873</td>
<td>0.441</td>
<td>0.635</td>
<td>8.03x10^{-6}</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs12931046</td>
<td>30987144</td>
<td>SETD1A</td>
<td>Intronic</td>
<td>G/A</td>
<td>0.873</td>
<td>0.441</td>
<td>0.639</td>
<td>8.03x10^{-6}</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs13708</td>
<td>31000809</td>
<td>HSD3B7</td>
<td>Downstream</td>
<td>G/A</td>
<td>0.873</td>
<td>0.441</td>
<td>0.636</td>
<td>8.03x10^{-6}</td>
<td>Imputed</td>
</tr>
<tr>
<td>rs1870293</td>
<td>30970941</td>
<td>SETD1A</td>
<td>Intronic</td>
<td>T/C</td>
<td>0.873</td>
<td>0.441</td>
<td>0.642</td>
<td>8.03x10^{-6}</td>
<td>Imputed</td>
</tr>
</tbody>
</table>

*Physical positions are based on build 37 (NCBI) of the human genome.

*1000 Genomes, Europeans.

LOC101928762, RNA gene, affiliated with the ncRNA class; ZNF668, zinc finger protein 668 gene; HSD3B7, hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 7; STX1B, syntaxin 1B; SETD1A, SET domain containing 1A.
Table 4.6. Allele and genotype frequencies of rs4918797 (chr10) and rs71383768 (chr16) in the studied population.

<table>
<thead>
<tr>
<th></th>
<th>rs4918797</th>
<th></th>
<th>rs71383768</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>C/C</td>
<td>C/T</td>
<td>T/T</td>
</tr>
<tr>
<td>All</td>
<td>106</td>
<td>50.0%</td>
<td>32.1%</td>
<td>17.9%</td>
</tr>
<tr>
<td>Sensitive</td>
<td>55</td>
<td>18.2%</td>
<td>47.3%</td>
<td>34.5%</td>
</tr>
<tr>
<td>Resistant</td>
<td>51</td>
<td>84.3%</td>
<td>15.7%</td>
<td>0.0%</td>
</tr>
<tr>
<td>1000 Genomes</td>
<td>503</td>
<td>64.6%</td>
<td>31.0%</td>
<td>4.4%</td>
</tr>
<tr>
<td>(Europeans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.6. LD plot of CYP2C9 SNPs, CYP2C9*2 (rs1799853), CYP2C9*3 (rs1057910), and rs4918797. Gabriel et al.'s definition (2002) was used to define the LD block in Haploview 4.2. The SNP of interest rs4918797 was in linkage disequilibrium with CYP2C9*2 (rs1799853) (D'=1.0, r^2=0.345) and CYP2C9*3 (rs1057910) (D'=1.0, r^2=0.437).
Figure 4.7. Genome-wide Manhattan plot (A) and regional association plot (B) of warfarin sensitive patients versus warfarin resistant patients, after conditioning for rs4918797. (A) Individual $-\log_{10} P$-values are plotted against their genomic position by chromosome. The top red line at $5 \times 10^{-8}$ marks the genome-wide significant threshold, while the blue line at $10^{-5}$ marks the threshold for promising SNPs. The red dots depict imputed SNPs. (B) Chromosome 16 is shown. The target SNP is depicted with a purple diamond. Each coloured circle represents a SNP $P$-value. SNPs with missing LD information are shown in grey. The relative positions of gene mapping to the region of association are also shown.
Figure 4.8. Genome-wide Manhattan plot (A) and regional association plot (B) of warfarin sensitive patients versus warfarin resistant patients, after conditioning for rs71383768. (A) Individual $-\log_{10} P$-values are plotted against their genomic position by chromosome. The top red line at $5 \times 10^{-8}$ marks the genome-wide significant threshold, while the blue line at $10^{-5}$ marks the threshold for promising SNPs. The green dots depict genotyped SNPs, and the red dots depict imputed SNPs. (B) Chromosome 10 is shown. The target SNP is depicted with a purple diamond. Each coloured circle represents a SNP $P$-value, with the colour scale reflecting the extent of LD ($r^2$). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on build 37 (NCBI) of the human genome. The relative positions of gene mapping to the region of association are also shown.
Figure 4.9. Genome-wide Manhattan plot of warfarin sensitive patients versus warfarin resistant patients, after conditioning for both rs4918797 and rs71383768. Individual $-\log_{10} P$-values are plotted against their genomic position by chromosome. The top red line at $5 \times 10^{-8}$ marks the genome-wide significant threshold, while the blue line at $10^{-5}$ marks the threshold for promising SNPs.
4.3.3.2 Warfarin extreme cases versus healthy controls

The NBS healthy controls were recruited from different parts of England and North Wales. Genotyping of the NBS healthy controls was performed on a different genotyping platform to our cohort of patients. To reduce population stratification artefacts and genotyping platform bias (Price et al., 2006), PCA was included as a covariate in the association analyses described in the following two sections.

4.3.3.3 Warfarin sensitivity versus healthy controls

When genotypic data from 55 patients sensitive to warfarin (41.8% male) were compared to those from 2,501 NBS healthy controls (49.5% male, \(P=0.262\)), genome-wide significant signals were observed on chromosome 7, chromosome 10, and chromosome 16 (Figure 4.10).

On chromosome 7, the most significant SNP was an imputed SNP, rs10236552 (\(P=8.11\times10^{-24}\); Figure 4.11A), an A/G single-nucleotide intronic variant located in the gene encoding for lipoma high-mobility group proteinIsoform I-C (HMGIC) fusion partner-like 3 (LHFPL3), which has been implicated in adipocytic tumours (Kubo et al., 2010). The rs10236552 homozygous recessive GG genotype was not present in patients sensitive to warfarin and NBS healthy controls. However, the frequency of the rs10236552 heterozygous AG genotype was over 12-fold higher in the group of patients sensitive to warfarin (40.0%) than the NBS healthy controls (3.2%) (Table 4.7).

Looking at the Manhattan plot in Figure 4.10, this SNP is clustered with two other imputed SNPs in isolation with very low \(P\)-values. Comparing the MAF
of this SNP in the NBS healthy controls to that of the European healthy volunteers from the 1000 Genomes project (0.016 versus 0.018, Table 4.7), the MAFs were found to be highly similar, suggesting that the genome-wide significant association was not due to an imputation error.

On chromosome 10, over 300 SNPs were genome-wide significant, and majority of them are located in the CYP2C9 gene (Figure 4.11B). The strongest signal was exhibited by an imputed SNP, rs1057911, an A/T single-nucleotide synonymous variant in the CYP2C9 gene \( (P=2.37 \times 10^{-18}) \). Interestingly, rs1057911 is in complete linkage disequilibrium with the genotyped SNP, CYP2C9*3 rs1057910 \( (D'=1.0, r^2=1.0; \text{Figure 4.12A}) \), which also attained genome-wide significance \( (P=2.47 \times 10^{-18}) \). The genotype frequencies of the homozygous recessive alleles of both rs1057911 (TT) and CYP2C9*3 rs1057910 (CC) were significantly increased in patients sensitive to warfarin when compared to those in the healthy controls (16.4% versus 0.4%; Table 4.7). CYP2C9*2 rs1799853, however, showed a \( P \)-value below the suggestive threshold of \( 1 \times 10^{-5} \) \( (P=0.002) \), and was not in LD with rs1057911 \( (D'=0.69, r^2=0.012; \text{Figure 4.12A}) \).

On chromosome 16, four imputed SNPs showed genome-wide significance (Figure 4.11C), rs2359612 \( (P=3.61 \times 10^{-11}) \), VKORC1*1639 rs9923231 \( (P=3.61 \times 10^{-11}) \), VKORC1 1173 rs9934438 \( (P=3.61 \times 10^{-11}) \), and rs8050894 \( (P=4.58 \times 10^{-11}) \). rs2359612 and rs8050894 are intronic SNPs located in the VKORC1 gene and are in high LD with VKORC1*1639 and VKORC1 1173 \( (\text{rs2359612, } D'=1.0, r^2=1.0; \text{rs8050894, } D'=1.0, r^2=0.988; \text{Figure 4.12B}) \). The frequency of VKORC1*1639 recessive AA genotype was higher in the group of patients sensitive to warfarin (63.5%) than that in the
healthy controls (14.9%, Table 4.7), confirming previous findings that the minor A allele of \textit{VKORC1} -1639 is associated with warfarin sensitivity (Klein \textit{et al.}, 2009; Limdi \textit{et al.}, 2010; Rieder \textit{et al.}, 2005; Wadelius \textit{et al.}, 2005, 2007; Wang \textit{et al.}, 2008).
Figure 4.10. Genome-wide Manhattan plot of warfarin patients sensitive to warfarin (n=55) versus healthy controls from NBS (n=2,501). PCA was included as a covariate. Individual $-\log_{10} P$-values are plotted against their genomic position by chromosome. The top red line at $5 \times 10^{-8}$ marks the genome-wide significant threshold, while the blue line at $10^{-5}$ marks the threshold for promising SNPs. The green dots depict genotyped SNPs ($CYP2C9^*3$ rs1057910, and rs4086116 respectively), while the red dots depict imputed SNPs.
Figure 4.11. Regional association plots of chromosomes 7 (A), 10 (B) and 16 (C). Patients sensitive to warfarin (n=55) were compared with healthy controls from NBS (n=2,501). The target SNPs are depicted with a purple diamond. Each coloured circle represents a SNP $P$-value, with the colour scale reflecting the extent of LD ($r^2$). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on build 37 (NCBI) of the human genome. The relative positions of gene mapping to the region of association are also shown.
Table 4.7. Allele and genotype frequencies of rs10236552 (chr7), CYP2C9\*3 rs1057910 and rs1057911 (chr10), rs2359612, VKORC1 -1639 rs9923231, and rs8050894 (chr16) in patients sensitive to warfarin and healthy controls.

<table>
<thead>
<tr>
<th></th>
<th>rs10236552</th>
<th>rs1057910</th>
<th>rs1057911</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>A/A</td>
<td>A/G</td>
</tr>
<tr>
<td>All</td>
<td>2556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy controls</td>
<td>2501</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitive</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000 Genomes</td>
<td>503</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Europeans)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>rs2359612</th>
<th>rs9923231</th>
<th>rs8050894</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>G/G</td>
<td>G/A</td>
</tr>
<tr>
<td>All</td>
<td>2556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy controls</td>
<td>2501</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitive</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000 Genomes</td>
<td>503</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Europeans)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.12. LD plot of (A) CYP2C9 SNPs, *2 rs1799853, *3 rs1057910, and rs1057911, and (B) VKORC1 SNPs, rs2359612, rs8050894, VKORC1 1173 (rs9934438), and VKORC1 -1639 (rs9923231). Gabriel et al.’s definition (2002) was used to define the LD block in Haploview 4.2. (A) On chromosome 10, the SNP of interest rs1057911 was not in linkage disequilibrium with CYP2C9*2 rs1799853 ($D'=0.69$, $r^2=0.012$), but was in strong LD with CYP2C9*3 rs1057910 ($D'=1.0$, $r^2=1.0$). (B) On chromosome 16, the two SNPs of interest rs2359612 and rs8050894 were in LD with VKORC1 1173 (rs2359612, $D'=1.0$, $r^2=1.0$; rs8050894, $D'=1.0$, $r^2=0.988$) and VKORC1 -1639 (rs2359612, $D'=1.0$, $r^2=1.0$; rs8050894, $D'=1.0$, $r^2=0.988$). Red squares with no numerical display depict $r^2=1.0$. 
4.3.3.4 Warfarin resistance versus healthy controls

When the 51 patients resistant to warfarin were compared to the 2,501 healthy controls, the clinical variable, gender, was found significant in the univariate analysis (66.7% versus 49.5% male, \( P=0.015 \)), and was therefore included as a clinical covariate in the GWAS logistic regression analysis.

Genome-wide significance was observed on chromosome 7 with a cluster of eleven imputed SNPs located in the \( LHFPL3 \) gene (Figure 4.13). All eleven SNPs were in LD among each other (Figure 4.14A). The most significant SNP was rs10236552 (\( P=2.07 \times 10^{-15} \)), an identical SNP to that found with the warfarin sensitivity versus healthy controls comparison as discussed in section 4.3.3.3. Similarly, the rs10236552 homozygous recessive GG genotype was not present in patients resistant to warfarin and NBS healthy controls. Whilst the frequency of the rs10236552 heterozygous AG genotype was over 9-fold higher in the group of patients resistant to warfarin (30.6%) than the NBS healthy controls (3.2%) (Table 4.8).

Suggestive genome-wide significance was observed with SNPs on chromosomes 6 and 18 (Figure 4.14B).

On chromosome 6, the top hit was an imputed SNP, rs114213056 (\( P=2.91 \times 10^{-6} \)), a C/T single-nucleotide intronic variant located in the \( MIR6873 \) (microRNA 6873) gene. The homozygous recessive TT genotype of rs114213056 was observed in a higher frequency in patients resistant to warfarin (9.8%) than the NBS healthy controls (2.1%) (Table 4.8).

On chromosome 18, four imputed SNPs (rs10163900, rs76455916, rs77118150, and rs79434376) showed suggestive genome-wide significance, all with \( P=7.82 \times 10^{-6} \) (Figure 4.14C). These four SNPs exhibited high LD with
each other (Figure 4.14C) and are all intronic variants located in the *PIGN* (phosphatidylinositol glycan anchor biosynthesis, class N) gene. The rs10163900 homozygous recessive genotype CC was not present in patients resistant to warfarin and NBS healthy controls. However, the frequency of the rs10163900 heterozygous CT genotype was over 6-fold higher in the group of patients resistant to warfarin (13.3%) than the NBS healthy controls (2.0%) (Table 4.8).
Figure 4.13. Genome-wide Manhattan plot of patients resistant to warfarin (n=51) versus healthy controls from NBS (n=2,501). PCA and gender were included as covariates. Individual $-\log_{10} P$-values are plotted against their genomic position by chromosome. The top red line at $5\times10^{-8}$ marks the genome-wide significant threshold, while the blue line at $10^{-5}$ marks the threshold for promising SNPs. The red dots depict imputed SNPs.
Figure 4.14. Regional association plots of chromosomes 7 (A), 6 (B) and 18 (C). Patients resistant to warfarin (n=51) were compared with healthy controls from NBS (n=2,501). The target SNPs are depicted with a purple diamond. Each coloured circle represents a SNP $P$-value, with the colour scale reflecting the extent of LD ($r^2$). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with a light blue line. Physical positions are based on build 37 (NCBI) of the human genome. The relative positions of gene mapping to the region of association are also shown.
Table 4.8. Allele and genotype frequencies of rs114213056 (chr6), rs10236552 (chr7), and rs10163900 (chr18) in patients resistant to warfarin and healthy controls.

<table>
<thead>
<tr>
<th></th>
<th>rs114213056</th>
<th></th>
<th>rs10236552</th>
<th></th>
<th>rs10163900</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>C/C</td>
<td>C/T</td>
<td>T/T</td>
<td>MAF</td>
<td>A/A</td>
</tr>
<tr>
<td>All</td>
<td>2556</td>
<td>73.7%</td>
<td>24.2%</td>
<td>2.1%</td>
<td>0.142</td>
<td>95.5%</td>
</tr>
<tr>
<td>Healthy controls</td>
<td>2501</td>
<td>74.3%</td>
<td>23.8%</td>
<td>2.0%</td>
<td>0.139</td>
<td>96.8%</td>
</tr>
<tr>
<td>Resistant</td>
<td>51</td>
<td>49.0%</td>
<td>41.2%</td>
<td>9.8%</td>
<td>0.304</td>
<td>69.4%</td>
</tr>
<tr>
<td>1000 Genomes (Europeans)</td>
<td>503</td>
<td>76.9%</td>
<td>21.5%</td>
<td>1.6%</td>
<td>0.123</td>
<td>96.4%</td>
</tr>
</tbody>
</table>
4.4 DISCUSSION

Genome-wide association studies offer a powerful approach to understand the genetic basis of drug response. Knowledge of additional polymorphisms affecting warfarin dose requirement in patients sensitive and resistant to would be extremely valuable in improving the safety and efficacy of warfarin therapy.

To date, we are the first to report a GWAS in patients requiring very high or low doses of warfarin.

Two intergenic variants, rs4918797 on chromosome 10 and rs71383768 on chromosome 16, exhibited suggestive genome-wide significant associations when warfarin sensitive patients were compared to patients with warfarin resistance. Interestingly, the minor alleles of these two SNPs were associated with warfarin sensitivity and conditional analyses suggest that the associations seen with rs4918797 on chromosome 10 and rs71383768 on chromosome 16 are correlated. In addition, rs4918797 was found to be in linkage disequilibrium with both CYP2C9*2 and *3, although CYP2C9*2 and *3 showed P-values below the threshold for suggestive associations. Our results suggest that rs4918797 could be a causal variant of warfarin sensitivity.

When patients sensitive to warfarin were compared to NBS healthy controls, genome-wide significant associations were observed with a synonymous variant in CYP2C9, rs1057911, which is in strong linkage disequilibrium with CYP2C9*3, and polymorphisms in the VKORC1 gene including VKORC1 -1639 and VKORC1 1173. Our results suggest that no other common variants in chromosomes 10 and 16 other than those in
CHAPTER 4

CYP2C9 and VKORC1, the major genetic determinants of warfarin dose, are associated with reduced warfarin dose requirements.

When warfarin resistant subjects were compared to NBS healthy controls, suggestive genome-wide significant association was observed with rs114213056 on chromosome 6, an intronic variant located in the MIR6873 gene. The function of this miRNA is currently unknown and further work is required to elucidate its role in warfarin resistance.

We have also identified four other variants on chromosome 18 (rs10163900, rs76455916, rs77118150, and rs79434376) which showed suggestive genome-wide significance when patients resistant to warfarin were compared to NBS healthy controls. These are intronic polymorphisms located in the PIGN gene. PIGN is involved in the glycosylphosphatidylinositol anchor-synthesis pathway (Maeda and Kinoshita, 2011). Mutations in this gene were reported to cause multiple congenital anomalies and severe neurological impairment (Maydan et al., 2011; Ohba et al., 2014). However, the role of this gene in blood coagulation is not known.

Interestingly, an intronic SNP located in the LHFPL3 gene on chromosome 7, rs10236552, showed genome-wide significant association in both warfarin sensitive versus NBS healthy controls and warfarin resistance versus NBS healthy controls comparisons. The rs10236552 heterozygous AG genotype was found to be present at a higher frequency in both our cohorts of patients with extreme warfarin phenotypes. Due to the isolation of the SNPs cluster as observed in the Manhattan plots, it is possible that this SNP effect is population or disease specific rather than drug/phenotypic specific. To verify this hypothesis, our patients showing sensitivity and resistance to warfarin will
need to be compared to patients requiring usual warfarin maintenance doses ranging between >1.5 mg per day to <10 mg per day.

One caveat of our study is the relatively small sample size. As such, our study may be underpowered to detect genome-wide significant differences for the numerous polymorphisms tested. Given that majority of the SNPs in our findings were imputed variants, genotyping will need to be performed to assess the validity of the association results to ensure that they are not imputation artefacts. Furthermore, our results need to be validated in a replication cohort to confirm the associations we have found.

In addition, to gain further insight into the biologic networks underlying the different warfarin extreme phenotypes, genes and pathway analyses need to be performed as potential combined effects of SNPs contributing to outcomes may be missed through conventional single marker associations.
CHAPTER 5

Final Discussion
## CONTENTS

### 5.1  Clinical implementation of pharmacogenetic testing ........... 190

- 5.1.1 Evaluation of effectiveness of genotype-guided warfarin dosing ................................................................. 191
- 5.1.2 Novel oral anticoagulants ................................................................. 194
- 5.1.3 Cost-effectiveness of oral anticoagulant therapy ................. 197

### 5.2  Our findings and future work ......................................................... 198

- 5.2.1 Pharmacogenetic-based warfarin dosing algorithm in Southern Italians ................................................................. 198
- 5.2.2 Circulating miRNAs as novel biomarkers of response to warfarin ........................................................................... 199
- 5.2.3 Warfarin Extreme Phenotypes: A Genome-Wide Association Study ........................................................................... 200

### 5.3  Conclusions .................................................................................... 202
5.1 Clinical implementation of pharmacogenetic testing

Although the post-genomic era is characterised by high-throughput genotyping technologies generating large amounts of genetic information, the clinical implementation of genomic data is lagging behind. The emergence of pharmacogenetic-guided treatment options that individualize therapy has enormous potential not only to offer safe and effective treatments, but also to lower overall health care costs through the prevention of adverse drug reactions that often result in repeated clinic visits or hospitalizations.

Over the past decade, there have been increasing efforts by regulatory agencies to survey, assess and document the contribution of pharmacogenetics in the inter-individual variability to drug response (Johnson and Cavallari, 2013; Lesko and Zineh, 2010; Zineh and Pacanowski, 2011). Currently, there are over 110 FDA-approved drugs with pharmacogenetic information in their labelling, (http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm), and warfarin is not an exception. In addition to regulatory agencies, the Clinical Pharmacogenetics Implementation Consortium (CPIC) of the National institute of Health’s Pharmacogenomics Research Network (http://www.pgrn.org), and PharmGKB aim to promote the widespread adoption of pharmacogenetic clinical application by the clinicians, through providing simplified, updated, freely and widely accessible guidelines and drug labels.

However, pre-emptive genetic testing has yet to be introduced into clinical practice (Prainsack and Vayena, 2013). The main barriers to the clinical implementation of pharmacogenetic testing include lack of knowledge among providers and interpretation of the results, financial support (such as
genotyping cost reimbursement), logistical facilities and robust evidence from large and well-conducted clinical trials (Johnson and Cavallari, 2013; Lesko and Zineh, 2010; Turner and Pirmohamed, 2014).

Indeed, prolonged and rather inconclusive controversies over the clinical utility and cost-effectiveness of pharmacogenetic-testing for warfarin have prevented its clinical implementation, despite the compelling evidence for its utility (Frueh, 2012).

5.1.1 Evaluation of effectiveness of genotype-guided warfarin dosing

Several non-randomized prospective studies aimed to evaluate pharmacogenetic-guided warfarin dosing have been published (Anderson et al., 2007; Epstein et al., 2010; Gong et al., 2011b; Lenzini et al., 2010; Wen et al., 2008). However, none of these studies was able to provide convincing evidence about the clinical significance of genotyping, either because of the small sample size or a non-randomized comparison (Verhoef et al., 2014).

At the end of 2013, results from the Clarification of Optimal Anticoagulation through Genetics (COAG), and the European Pharmacogenetics of Anticoagulant Therapy (EU-PACT) trials were simultaneously published in the New England Journal of Medicine (Kimmel et al., 2013; Pirmohamed et al., 2013). The COAG trial, conducted in the U.S., was a multi-centre, double-blinded RCT comparing genotype guided warfarin dosing with a clinical dosing algorithm. For genotype-guided dosing, the COAG researchers utilised the dose-initiation algorithm by (Gage et al., 2008) and a dose revision algorithm by Lenzini et al. (2010). The EU-PACT trial was
a single-blinded, multi-centre RCT conducted in the UK and Sweden, comparing genotype guided warfarin dosing to standard care. Importantly, because of the pharmacokinetics of warfarin, the trial also used loading doses at initiation. For genotype-guided dosing, the EU-PACT team used a modified version of the IWPC algorithm during warfarin therapy initiation (to incorporate loading doses) and the same dose revision algorithm as that used in the COAG trial (Klein et al., 2009; Lenzini et al., 2010).

Both trials evaluated the same outcome, the effect of genotype-guided dosing strategy on the percentage of time in therapeutic INR range (TTR), albeit for different intervals (4 weeks [COAG] vs 12 weeks [EU-PACT]). The COAG authors did not find any between-groups differences in the mean TTR after 4 weeks of therapy (Kimmel et al., 2013). In contrast, results of the EU-PACT trial showed a significantly higher percentage of patients (7%) with TTR in the genotype-guided arm after 3 months of treatment. Reasons for the divergent results observed in these two RCT trials are four-fold:

(i) In the EU-PACT trial, the genotype group was compared with standard dosing methods used in clinical practice, while the COAG trial used clinical variables incorporated in an algorithm for the clinically-based dosing strategy. This is suggested to have let to possible more precise control in the INR values (Zineh et al., 2013), although this is purely speculative and difficult to explain given that the overall contribution of clinical factors is significantly less than genetic factors.

(ii) In the EU-PACT trial, genotype results were available within two hours, allowing all day one dosing to be based on genotype. In the
COAG trial, only 45% of patients had genotype results on day one, 99% of subjects had genetic information only by day three. Thus the different timing of genotyping results availability could have altered the results of the outcome (Baranova et al., 2014; Johnson and Cavallari, 2015).

(iii) In the COAG trial, one-third of the patients were African-Americans while patients in the EU-PACT trial were all Caucasians. Considering that the frequency of the VKORC1 and CYP2C9 alleles varies between different ethnic groups and, in African Americans, other genetic variants such as CYP2C9 *5, *6, *8, and *11 (Hernandez et al., 2014; Suarez-Kurtz and Botton, 2013) which were not assessed may contribute to the variability in patient response to warfarin.

(iv) It is also postulated that regional differences in anticoagulation care between the U.S. and Europe may have also played a role in the difference in the observed results (Baranova et al., 2014; Johnson and Cavallari, 2015).

Several meta-analyses of the largest RCTs have been conducted to provide more evidence on the effect of the genotype-guided warfarin dosing on thromboembolic and haemorrhagic complications (Franchini et al., 2014; Goulding et al., 2015; Liao et al., 2015; Stergiopoulos and Brown, 2014; Tang et al., 2014). However, results showed similar conflicting results. While some authors found a significant reduction in warfarin-related bleeding events with genotype-based dosing approach compared with clinical dosing approach (Franchini et al., 2014; Goulding et al., 2015; Tang et al., 2014), others failed to find any significant difference (Liao et al., 2015; Stergiopoulos and Brown,
2014), suggesting that pooled data from existing trials might be insufficient to detect significant differences in clinically relevant endpoints.

Even though no one disputes the influence of genetics on warfarin dosing, clinical benefits of pharmacogenetic-based warfarin dosing algorithms is still debatable. Therefore, several clinical trials are currently ongoing to better clarify this issue.

### 5.1.2 Novel oral anticoagulants

Recently, several novel oral anticoagulants (NOACs), also called direct oral anticoagulants (DOACs) have been introduced for the treatment and prevention of thrombosis (Bauer, 2013). Dabigatran is a selective, direct thrombin inhibitor (Fareed et al., 2012; Saraf et al., 2014), while rivaroxaban, apixaban and edoxaban are reversible factor Xa inhibitors (Bounamenteaux and Camm, 2014; Perzborn et al., 2011; Wong et al., 2008, 2011) (Figure 5.1).

NOACs have significant advantages over warfarin including (i) a more predictable pharmacokinetic/pharmacodynamic profile, (ii) fixed dosing without the need for routine coagulation monitoring, (iii) relatively rapid onset and offset of action which helps to facilitate management of patients requiring surgery or interventions, and (iv) reduced susceptibility to food and drug interactions.

However, NOACs still have some drawbacks preventing their wide adoption. Currently, it is not mandatory to check the degree of anticoagulation via laboratory testing (Samama and Guinet, 2011); as such, patients’ compliance cannot be easily assessed. NOACs have been shown to be
associated with potential increased risk of gastrointestinal bleeding and myocardial infarction, suggesting the need of some form of anticoagulation monitoring (Connolly et al., 2010; Granger et al., 2011; Miller et al., 2012; Patel et al., 2011; Ruff et al., 2014). As NOACs are metabolised by liver enzymes and eliminated as active drugs in the kidneys, they are not recommended in patients with severe hepatic failure and renal dysfunction (Douxfils et al., 2014; Wang and Bajorek, 2014). Current fixed dosing of NOACs are not suitable for patients with mechanical heart valves (Eikelboom et al., 2013). In addition, the suitability of NOACs in paediatric patients have not been tested, limiting their prescription in this population. Although several reversal agents in the event of bleeding are in development, currently the monoclonal antibody idarucizumab is the only antidote that has been approved for dabigatran (Hu et al., 2016; Mo and Yam, 2015; Pollack et al., 2015).

Given the limited experience with the NOACs, their limited range of indications, and their associated high costs, the characteristics of each oral anticoagulant must be considered carefully to select the best agent to provide the optimal risk/benefit profile in patients.
Figure 5.1. Point of action of novel oral anticoagulants in the coagulation cascade. Direct Xa and thrombin (IIa) inhibitors interact with their respective factors (Connolly et al., 2009; Eerenberg et al., 2011; Giugliano et al., 2013; Granger et al., 2011; Turpie et al., 2009). Green arrows depict activation, while red lines depict inhibition. HMWK, high-molecular-weight kininogen; PK, prekallikrein; TF, tissue factor; TFPI, tissue factor pathway inhibitor; APC, activated protein C.
5.1.3 Cost-effectiveness of oral anticoagulant therapy

Numerous studies assessing the cost-effectiveness of pharmacogenetic-guided warfarin dosing compared to clinical dosing have been published (Eckman et al., 2009; Leey et al., 2009; Meckley et al., 2010; Patrick et al., 2009; Verhoef et al., 2010; You et al., 2009), but with limited conclusions about their economic value, principally because of the uncertainty in the clinical evidence gathered. Several factors affect the cost-effectiveness of pharmacogenetic-guided warfarin dosing including the population where it is tested, the indication, the age of the patients and the cost of the pharmacogenetic test as well as how often it will be used (Verhoef et al., 2012). In addition, many differences exist between European countries in the organization of anticoagulation services, and the costs of anticoagulant therapy (Verhoef et al., 2012). Therefore, it is likely that the cost-effectiveness of pharmacogenetic-guided warfarin dosing will vary considerably among countries.

Recent studies examined the cost-effectiveness between genotype-guided warfarin dosing and treatment with NOACs (Pink et al., 2014; You, 2014; You et al., 2012), showing that the cost-effectiveness of genotype-guided warfarin dosing is largely dependent on the quality of warfarin anticoagulation control, i.e. time within therapeutic range.

A simulation approach was recently conducted to assess the cost-effectiveness of genotype-guided dosing of warfarin and NOACs whereby the decision of oral anticoagulation therapy was based on the CYP2C9 and VKORC1 genotypes of patients (You, 2014). In the simulation study, patients carrying the wild-type genotypes for CYP2C9 and VKORC1 were dosed on
warfarin using a pharmacogenetic algorithm while patients carrying the variant alleles were assumed to be susceptible to over-anticoagulation and were therefore treated with a NOAC. Using this stratified approach, pharmacogenetic dosing was shown to be highly cost-effective. Definitive evidence from clinical trials, however, will be needed to confirm these simulations results.

5.2 Our findings and future work

5.2.1 Pharmacogenetic-based warfarin dosing algorithm in Southern Italians

In chapter 2, contrary to previous studies conducted in patients from Central and Northern Italy (D’Andrea et al., 2005; Borgiani et al., 2007, 2009; Zambon et al., 2011), we did not find any association between VKORC1 1173C>T and CYP4F2 1297G>A polymorphisms with warfarin stable dose in our cohort of 266 patients from the Campania Region (Southern Italy). We developed a pharmacogenetic warfarin dosing algorithm incorporating six variables (age, BSA, amiodarone status, CYP2C9*2, CYP2C9*3, and VKORC1 -1639G>A). Our algorithm allowed the accurate prediction of warfarin maintenance dose in 44% of patients in our cohort, with a mean absolute error of 7.41 mg/week. The prediction accuracy of our pharmacogenetic algorithm was similar to that of the IWPC algorithm and superior to the three published Italian pharmacogenetic algorithms derived from patients in Northern and Central Italy (Borgiani et al., 2009; Cini et al., 2012; Zambon et al., 2011). Our results, in accordance with data from two
GWAS conducted in Italian patients (Di Gaetano et al., 2012; Fiorito et al., 2015), confirmed that patients from Southern Italy should be considered genetically different to patients from Northern and Central Italy.

A prospective observational multicentre study funded by the Italian Medication Agency (AIFA) for estimating the appropriate initial dose of warfarin in elderly patients (≥65 years) from Southern Italy with heart valves or non-valvular atrial fibrillation and at least one comorbid condition, is currently ongoing (EUDRACT number: 2012-002578-30; AIFA code: FARM9JNT9Y). Elderly patients constitute a population highly vulnerable to bleeding during treatment with warfarin, mainly because of comorbidity, consumption of interacting drugs or reduced compliance (Fihn et al., 1996; Mansur et al., 2012; Wieloch et al., 2011). Further work will be undertaken to investigate whether the use of a pharmacogenetic-based dosing algorithm could better predict warfarin dose requirements in this fragile population, and reduce haemorrhagic and thrombotic events.

5.2.2 Circulating miRNAs as novel biomarkers of response to warfarin

Recently, great progress has been made in the field of “omics” technologies, where RNA, protein and metabolome profiling along with genomic information are being integrated to create a more comprehensive strategy for personalized medicine (Chen et al., 2012).

In chapter 3, we conducted miRNA profiling on baseline serum from patients who have achieved warfarin stable dose (n=10) and those who did not achieve warfarin stability (n=10), using the Affymetrix miRNA array.
Although previous studies suggested a role for miR-133a in warfarin response (Ciccacci et al., 2015; Pérez-Andreu et al., 2012; Shomron, 2010), we did not find any association with levels of circulating miR-133a in our pilot study. Interestingly, circulating levels of a novel miRNA, miR-548a-3p, were observed to be significantly higher in patients who did not achieve warfarin stable dose. *In silico* analyses showed that several target genes of miR-548a-3p are involved in the coagulation pathway, suggesting that this circulating miRNA could play a role in warfarin dose variability, acting as a potential biomarker for warfarin treatment response.

To date, we are the first group who has explored the effect of circulating miRNAs on warfarin therapy response. We are aware of the small sample size of our pilot study and will therefore be validating and replicating the effects of circulating miR-548a-3p and miR-133a on warfarin response in a larger cohort of prospectively recruited patients initiated onto warfarin therapy (n=980). We have optimised a real-time quantitative PCR method to assess miRNA expression levels in patients’ sera in chapter 3 and will be using this optimised qPCR method to validate and replicate our initial findings. Further work will include incorporating validated miRNAs into warfarin dosing algorithms. The biological effects of miRNAs on target genes will also be investigated.

### 5.2.3 Warfarin Extreme Phenotypes: A Genome-Wide Association Study

As the candidate gene approach focuses on potential genes with functional relevance, effects of unknown genes could not be determined. An
unbiased approach like GWAS is hypothesis-free and allows the investigation of common SNPs across the entire human genome (Daly, 2010; Harper and Topol, 2012; Ni et al., 2013).

Our GWAS data in chapter 4 demonstrated that in addition to the common variants in CYP2C9 and VKORC1, rs4918797, an intergenic SNP on chromosome 10, was associated with warfarin sensitivity.

Suggestive association with warfarin resistance was observed with intronic SNPs in MIR6873 (rs114213056), and PIGN (rs10163900, rs76455916, rs77118150, and rs79434376). How these genes confer warfarin resistance need further evaluation.

An adequate sample size is very important to detect genome-wide significance for polymorphisms that modestly alter warfarin response (Takeuchi et al., 2009). Further validation of our findings in a replication cohort is needed. An international multi-centre collaboration would be a good strategy to increase the sample size. Such collaboration would permit a more accurate representation of the distribution of genotypes amongst patients sensitive or resistant to warfarin, and may help in determining the relative contributions of both non-genetic and genetic factors to warfarin response in these patients.

Studies have reported associations of rare mutations in patients who require unusually high or low doses of warfarin, but focus was largely on two selected genes, CYP2C9 and VKORC1 (Bodin et al., 2005a, 2005b; Ciccacci et al., 2015; D’Ambrosio et al., 2007; Di Fusco et al., 2013; Shuen et al., 2012). Rare variants in other genes may also play a role in warfarin sensitivity and warfarin resistance. Exome-Chip analysis has been performed in our cohorts of warfarin sensitive and warfarin resistant patients. Analysis is currently
underway to identify underlying rare mutations, which could hopefully aid the clinical management of patients with extreme warfarin dose requirements.

5.3 Conclusions

Warfarin has been the mainstay of oral anticoagulation across multiple clinical indications over the past 60 years. After the introduction of NOACs in 2009, a rapid adoption of novel anticoagulants into clinical practice has been observed (Desai et al., 2014; Luger et al., 2014). Although the possibility of fixed-dose NOACs regimens without the need for regular coagulation monitoring may appeal to many physicians and patients alike, there are also patients such as those with mechanical heart valves who will benefit from the individual dosing and close monitoring that are prerequisites for warfarin treatment. Currently, the therapy decision between warfarin and NOACs is determined by several factors including age, indication, renal and hepatic function, risk/benefit profile, comorbidities and concomitant medications. In the long term, it is possible that genetic information of patients will be available in their electronic records (Maitland-van der Zee et al., 2014). We could therefore use genetic data as we currently use routine biochemical data. Indeed, several clinics in the U.S. are already using pharmacogenetic data in real-life setting (Nutescu et al., 2013). Furthermore, a stratified approach as proposed by You (2014) whereby the choice of anticoagulant prescription is made on the basis of pharmacogenetic information could help identifying the best drug for each patient.
In this scenario, further investigations on warfarin pharmacogenomics are still worthwhile as they will provide valuable lessons to personalise anticoagulant therapy. This thesis demonstrates that non-coding RNA and additional genetic variants may be involved in determining warfarin dosing variability, and they may help improving the accuracy of pharmacogenetic-guided warfarin dosing.


BIBLIOGRAPHY


BIBLIOGRAPHY


APPENDIX
Appendix 1.1a. Characteristics of the patients according to CYP2C9*2, CYP2C9*3, and CYP4F2 1297G>A genotypes.

<table>
<thead>
<tr>
<th>Indication</th>
<th>CYP2C9*2</th>
<th>CYP2C9*3</th>
<th>CYP4F2 1297G&gt;A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*1/*1 (197)</td>
<td>*1/*2 (63)</td>
<td>*2/*2 (6)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>66.52±11.50</td>
<td>68.10±10.54</td>
<td>77.50±8.64</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>101/96</td>
<td>41/22</td>
<td>5/1</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>26.95±4.50</td>
<td>26.81±3.73</td>
<td>27.60±1.14</td>
</tr>
</tbody>
</table>

| Indication                  | *1/*1 (224)       | *1/*3 (38)        | *3/*3 (4)          | G/G (121) | G/A (123) | A/A (22) |
| Atrial fibrillation         | 87                | 24                | 2                  | 98        | 13        | 2        | 55        | 47        | 11        |
| Dilatative cardiomyopathy   | 70                | 24                | 4                  | 83        | 14        | 1        | 40        | 50        | 8         |
| Deep venous thrombosis      | 16                | 6                 | 0                  | 17        | 4         | 1        | 7         | 13        | 2         |
| Pulmonary embolism          | 14                | 5                 | 0                  | 13        | 6         | 0        | 10        | 8         | 1         |
| Other                       | 5                 | 2                 | 0                  | 6         | 1         | 0        | 4         | 3         | 0         |
| INR target (range)          | 2.5 (2.0-3.0)     | 38                 | 5                  | 141       | 27        | 3        | 79        | 80        | 12        |
|                            | 3.0 (2.5-3.5)     | 25                 | 1                  | 83        | 11        | 1        | 42        | 44        | 9         |

Continuous variables are expressed as mean ± standard deviation (SD) and categorical variables as number of subjects.
Appendix 1.1b. Characteristics of the patients according to CYP2C9*2, CYP2C9*3, and CYP4F2 1297G>A genotypes.

<table>
<thead>
<tr>
<th></th>
<th>CYP2C9*2</th>
<th>CYP2C9*3</th>
<th>CYP4F2 1297G&gt;A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*1/*1 (197)</td>
<td>*1/*2 (63)</td>
<td>*2/*2 (6)</td>
</tr>
<tr>
<td>Smoking</td>
<td>16</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Dyslipidaemia</td>
<td>124</td>
<td>43</td>
<td>6</td>
</tr>
<tr>
<td>Hypertension</td>
<td>119</td>
<td>43</td>
<td>5</td>
</tr>
<tr>
<td>Diabetes</td>
<td>31</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Liver disease</td>
<td>29</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Kidney failure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild renal impairment</td>
<td>8</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Severe renal impairment</td>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Concomitant drug</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>assumption</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enzyme inducer status</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Phenyltoin</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Phenytoin</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Phenobarbital</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Enzyme inhibitor status</td>
<td>28</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>Amiodarone</td>
<td>15</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Omeprazole</td>
<td>9</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>8</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Values represent the number of subjects.
### Appendix 1.2a. Characteristics of the patients according to VKORC1 -1639G>A, VKORC1 1173C>T, and VKORC1 3730G>A genotypes.

<table>
<thead>
<tr>
<th></th>
<th>VKORC1 -1639G&gt;A</th>
<th></th>
<th>VKORC1 1173C&gt;T</th>
<th></th>
<th>VKORC1 3730G&gt;A</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G/G (69)</td>
<td>G/A (131)</td>
<td>A/A (66)</td>
<td>C/C (114)</td>
<td>C/T (116)</td>
<td>T/T (36)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>66.86±12.47</td>
<td>66.49±10.79</td>
<td>68.95±10.85</td>
<td>67.89±11.82</td>
<td>67.51±10.21</td>
<td>64.35±12.75</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>41/28</td>
<td>64/67</td>
<td>42/24</td>
<td>71/43</td>
<td>56/60</td>
<td>20/16</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>27.44±3.83</td>
<td>27.09±4.29</td>
<td>26.25±4.54</td>
<td>26.54±3.91</td>
<td>27.23±4.47</td>
<td>27.13±4.73</td>
</tr>
<tr>
<td>Stable dose (mg/wk)</td>
<td>36.90±14.20</td>
<td>29.12±11.79</td>
<td>18.82±7.99</td>
<td>29.05±14.73</td>
<td>28.05±11.69</td>
<td>26.82±12.36</td>
</tr>
<tr>
<td>Indication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac valve replacement</td>
<td>25</td>
<td>63</td>
<td>25</td>
<td>42</td>
<td>49</td>
<td>22</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>30</td>
<td>43</td>
<td>25</td>
<td>45</td>
<td>46</td>
<td>7</td>
</tr>
<tr>
<td>Dilative cardiomyopathy</td>
<td>7</td>
<td>11</td>
<td>4</td>
<td>9</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Deep venous thrombosis</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>INR target (range)</td>
<td>2.5 (2.0-3.0)</td>
<td>47</td>
<td>88</td>
<td>36</td>
<td>73</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>3.0 (2.5-3.5)</td>
<td>22</td>
<td>43</td>
<td>30</td>
<td>41</td>
<td>36</td>
</tr>
</tbody>
</table>

Continuous variables are expressed as mean ± standard deviation (SD) and categorical variables as number of subjects.
### Appendix 1.2b. Characteristics of the patients according to VKORC1 -1639G>A, VKORC1 1173C>T, and VKORC1 3730G>A genotypes.

<table>
<thead>
<tr>
<th></th>
<th>VKORC1 -1639G&gt;A</th>
<th>VKORC1 1173C&gt;T</th>
<th>VKORC1 3730G&gt;A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G/G</td>
<td>G/A</td>
<td>A/A</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(69)</td>
<td>(131)</td>
<td>(66)</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>7</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Dyslipidaemia</td>
<td>44</td>
<td>87</td>
<td>42</td>
</tr>
<tr>
<td>Hypertension</td>
<td>44</td>
<td>80</td>
<td>43</td>
</tr>
<tr>
<td>Diabetes</td>
<td>13</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>Liver disease</td>
<td>11</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>Kidney failure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild renal impairment</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Severe renal impairment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concomitant drug assumption</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enzyme inducer status</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Phenytoin</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Phenobarbital</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Enzyme inhibitor status</td>
<td>11</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Amiodarone</td>
<td>8</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Omeprazole</td>
<td>3</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Simvastatin</td>
<td>1</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

Values represent the number of subjects.