CHAPTER IV

AN INVESTIGATION OF GENETIC DETERMINANTS OF WARFARIN DOSE REQUIREMENT IN 201 SWEDISH PATIENTS (UPPSALA COHORT)

4.1 PATIENT RECRUITMENT

Collection and genetics analysis of subjects in the Uppsala cohort was approved by the local Ethics Committee in Uppsala, Sweden. After informed consent, Dr Mia Wadelius and colleagues (Uppsala University) collected blood samples from 201 patients who were treated with warfarin at the anticoagulation clinic at the Uppsala University Hospital in 2000. These patients were aged from 28 to 88 years old. At the time of recruitment, patients had been treated with warfarin for at least 2 months, and the frequency each patient visited the clinic varied from every week to every month, depending on medications. Their treating period varied from 2.4 months to 26 years and the median duration of treatment was 2 years.

Medical information for each patient was recorded at six consecutive visits in the clinic. For each visit, the administered dose and corresponding PT INR value was recorded in the database. The prescribed warfarin dose for these patients varied from 4.5 to 77.25 mg/week. Meanwhile, information about age, gender, body weight, and any other diseases and indication for concurrent treatment was extracted from the patients' medical records (Table 4.1). 113 (56%) patients in this study had atrial fibrillation. The second largest medical indication was those patients with heart valve prosthesis (49 individuals, 24.4%). These patients were sorted into a separate group because a higher PT INR value is targeted. The average PT INR of majority patients (153 patients) was within 2 and 3 while 13 patients had the average INR below 2 and 35 patients had the average INR above 3. Among the total of 1005 patient-visits, 686 (68.3%) of measured INRs were between 2 and 3.

Most patients had other diseases including hypertension (78 patients), heart failure (51 patients), angina pectoris (35 patients) and type 2 diabetes (18 patients) and received

concurrent medications.

Table 4.1. Medical information statistics of patients in Uppsala cohort.

Characteristics		Patients	(%)	Effects on warfarin
Indication				
	Atrial fibrillation	113	56.2	
	Heart valve prosthesis	49	24.4	
	Deep vein thrombosis/pulmonary embolus	9	4.5	
	Cardiomyopathy	8	4.0	
	Transischemic attack	5	2.5	
Other diseases				
	Hypertension	78	38.8	
	Heart failure	51	25.4	
	Angina pectoris	35	17.4	
	Type 2 diabetes mellitus	18	9.0	
Interacting medication				
	Simvastatin	25	12.4	+
	Aspirin	21	10.4	+
	Paracetamol	18	9.0	+
	Amiodarone	9	4.5	+
	Disopyramide	7	3.5	+
	Dextropropoxyphene	7	3.5	+
	Propafenone	3	1.5	+
	Carbamazepine	3	1.5	-
	Non-steroidal anti-inflammatory drug	2	1.0	+
	Phenytoin	1	0.5	-
	Mianserin	1	0.5	-
Gender				
	Men	135	67.2	
	Women	66	32.8	
Age				
	Mean years (range)	66.9	22-88	

As to the concurrent medications, the drugs were catalogued as interacting if they had moderate or major interactions with warfarin according to the database MICROMEDEXs Healthcare Series (http://www.micromedex.com/). In the 201 Uppsala patients, a total of 107 concurrent medications were known to influence warfarin (Table 4.1). Therefore, patients were divided into three groups for further statistical analysis: four individuals were simultaneously taking drugs that lower the effect of warfarin, 74 patients received

medications that enhanced the effect of warfarin and the 123 patients were not on any medications known to affect the metabolism of warfarin.

Ethnic origin is also an important factor in genetic studies and should be accounted for in order to avoid false positive associations due to confounding effects (Devlin et al. 2001). The patients were mainly Caucasians. Of the 201 patients, 194 were of Swedish origin; four of other European descent and three from the Middle East.

4.2 UNIVARIATE ASSOCIATION ANALYSIS

As described in chapter III 783 SNPs (1146 SNPs attempted) were successfully genotyped in the 35 candidate genes selected to test for association to warfarin dose requirement. To best fit the linear model, the association between SNPs and average weekly dose was analysed using the original value, log of dose and square root of dose. Square root demonstrated better distribution in linearity. Univariate regression analysis was then performed and Table 4.2 shows all the nominally associated SNPs ($P \leq 0.05$) with square root of dose ranked by p-value. Only 379 common SNPs ($MAF \geq 0.05$) were included in the analyses. In Table 4.2, N is the number of successfully genotyped individuals for each SNP, whereas univariate R^2 indicates the ratio of inter-individual dose variation explained by that SNP in a regression model. The LD (r^2) between the SNP with the lowest P-value and any other in the gene or gene cluster (*CYP2C9*, *CYP2C19*, and *CYP2C18*) was also calculated and is shown in the last column (Table 4.2).

The 33 SNPs nominally associated with warfarin dose ($P \leq 0.05$) were found in *VKORC1, CYP2C9-CYP2C19-CYP2C18, PROC, EPHX1*, *CALU*, *PDIA2*, *GGCX*, *ORM1-2* and *APOE*. Notable, the two polymorphisms (rs429358 and rs7412) in *APOE* were used to discriminate haplotypes between APOE*E2, *E3 and *E4 allele, and the association of the dose and *APOE* is significant only when the haplotypes were assessed as E2 + E4 versus E3.

Table 4.2. Univariate analysis in regression model.

Gene	SNP	MAF	N	Univariate R^2	P-value	r^2 with best SNP
VKORC1	rs9923231	0.391	181	0.317	$1.91\times10^{-15**}$	-
	rs2359612	0.389	200	0.29	$2.30\times10^{-15**}$	0.968
	rs9934438	0.383	169	0.292	$3.59\times10^{-13**}$	1
	rs7294	0.384	188	0.208	$4.14\times10^{-10**}$	0.385
	rs4889490	0.446	199	0.16	$3.821\times10^{-8**}$	0.461
	rs4889537	0.372	199	0.142	$3.158\times10^{-7**}$	0.209
	rs4889599	0.366	194	0.124	$3.270\times10^{-6**}$	0.305
	rs8046978	0.214	197	0.047	0.00906	0.173
	rs11642603	0.093	192	0.027	0.02304	0.07
	rs11642466	0.103	195	0.025	0.02623	0.08
	rs7194347	0.343	197	0.032	0.04069	0.153
CYP2C9	rs1057910 (*3)	0.058	201	0.141	$2.784\times10^{-7**}$	-
	rs9332108	0.064	201	0.141	$2.784\times10^{-7**}$	0.89
	rs9325473	0.055	189	0.147	$3.753\times10^{-7**}$	0.908
	rs1057911	0.067	191	0.145	$4.218\times10^{-7**}$	0.89
	rs9332214	0.059	198	0.139	$4.654\times10^{-7**}$	0.878
	rs4917639	0.173	197	0.118	$4.944\times10^{-6**}$	0.276
	rs2860905	0.214	193	0.072	0.0008*	0.224
CYP2C19	rs3814637	0.059	195	0.106	0.00002**	0.838[a]
	rs17882687(*15)	0.08	183	0.044	0.00417*	0.395[a]
CYP2C18	rs7896133	0.056	193	0.074	0.00013	0.869[a]
PROC	rs2069919	0.372	182	0.09	0.00022*	-
	rs1799809	0.433	188	0.078	0.00055*	0.777
	rs2069901	0.441	177	0.072	0.00147*	0.785
	rs2069910	0.387	178	0.046	0.01678	0.414
APOE	rs429358+rs7412[b]	0.251	201	0.051	0.0057*	-
EPHX1	rs4653436	0.266	196	0.048	0.00848	-
CALU	rs11653	0.366	197	0.047	0.00944	-
	rs1006023	0.331	200	0.033	0.03789	0.865
	rs2307040	0.336	200	0.033	0.03811	0.867
	rs339054	0.461	195	0.032	0.04487	0.612
PDIA2	rs400037	0.232	183	0.020	0.02954	-
GGCX	rs12714145	0.408	198	0.034	0.0332	-
ORM1-2	rs1687390	0.062	149	0.026	0.04964	-

** Experiment-wise significance, significant p-value is 1.65E-04.

* Gene-wise significance, significant p-value depends on the effective number of test in each gene or gene cluster.

[a] Linkage disequilibrium with CYP2C9*3 (rs1057910)

[b] The two APOE SNPs are not significant individually, the significance derived from assessed as E2+E4 vs. E3

To minimise false positives it is important to correct p-values for performing multiple tests.

Bonferroni correction is the most frequent approach of correcting for multiple tests, but this

estimate is very conservative as it considers all tests as independent (Balding 2006); for example, tightly linked SNPs in terms of LD constitute one as opposed to multiple tests. The Bonferroni correction based on the effective number of independent traits (Bonferroni Meff) provides a simple way of estimation of significant p-value after multiple testing (Cheverud 2001; Li 2001; Nyholt 2004), which is close to that of the permutation procedure. However, multiple correction using permutation procedure is very computationally demanding. In the dose study, we therefore corrected the p-value using the Bonferroni Meff method. Since there is no evidence showing that the 35 selected genes have co-evolved, we assume that genes are independent of each other, and SNPs tightly linked to the tag SNPs are disregarded. The Meff estimation is obtained using Single Nucleotide Polymorphism Spectral Decomposition (SNPSpD, http://gump.qimr.edu.au/general/daleN/SNPSpD/). The sum of effective tests over all genes is 303 in this study and the experiment-wise significance is $P < 1.65E-04$. Meanwhile, the number of independent effective tests within each gene or gene cluster varied from 2 to 50, resulting in gene-wise corrected significance of $P < 0.025$ (two independent SNPs) to $P < 0.001$ (50 independent SNPs). After experiment-wise correction for multiple testing, only *VKORC1* (7 SNPs), *CYP2C9* (6 SNPs), *CYP2C18* (1 SNP) and *CYP2C19* (1 SNP) were associated with warfarin dose. *PROC* and *APOE*, however, passed the gene-wise correction but failed in the experiment-wise test (Table 4.2).

The two SNPs in *CYP2C19* and one SNP in *CYP2C18* (Table 4.2) are in strong linkage disequilibrium with *CYP2C9**3 allele (rs1057910); which is known to be associate with warfarin dose (Aithal et al. 1999; Higashi et al. 2002; Wadelius et al. 2005; Wadelius et al. 2004). The *VKORC1* and CYP2C cluster loci are discussed in detail in section 4.3 and 4.4.

Association analysis using 2- or 3-marker sliding-window haplotypes was also performed for

each candidate gene. In each sliding-window the haplotype was phased statistically according to alleles of each SNP. We found 13 genes to be nominally associated ($P < 0.05$) with warfarin dose, and Table 4.3 lists only the haplotypes that exhibited the lowest p-value in each candidate gene for association with the square root of dose. Compared to single marker analysis (Table 4.2), six additional genes, *PDIA6*, *F7*, *PDIA3*, *PROZ*, *F9* and *NR1I2*, were nominally associated. *EPHX1*, *CALU*, *PDIA2* and *ORM1-2* did not have a haplotype that reached nominal significance.

Haplotypes in *VKORC1*, *CYP2C9*, *CYP2C18* and *CYP2C19* were the most associated and showed experiment-wise significance as they did in single marker analysis. As with single marker analysis, *PROC* was also significant at the gene-wise level in haplotype analysis. Haplotypes giving experiment-wise significance in *CYP2C9*, *CYP2C18* and *CYP2C19* are associated with either *CYP2C9**2 or *CYP2C9**3 alleles. In *GGCX*, the CAA microsatellite (section 4.3.3) was taken into account with the SNP haplotype. Table 4.3 shows the result of two or three marker haplotype association in dose. The last column in Table 4.3 compares the p-value between single marker and haplotype analysis. Haplotypes in *VKORC1*, *PROC* and *ORM1-2* provide no additional information over single SNPs. The remaining ten genes in Table 4.3 show a slightly smaller p-value in haplotype analysis than in single marker analysis. The experiment-wise significant haplotypes in *CYP2C9*, *CYP2C18* and *CYP2C19* have one or more SNP tightly linked to either *CYP2C9**2, *CYP2C9**3 or both.

Table 4.3. Two or three marker haplotype association in dose variation.

Gene	Haplotype	P-value	smaller p-value
VKORC1	rs9934438-rs9923231	5.76E-15**	**No**
CYP2C9	rs9332214[a] -rs9332222[b]-rs2298037	4.86E-09**	Yes
CYP2C18	rs1926711-rs7919273[a]-rs10509675	3.47E-07**	Yes
CYP2C19	rs2860840-rs3814637[a]	2.08E-06**	Yes
PROC	rs2069919-rs2069921-rs973760	1.36E-03*	**No**
PDIA6	rs11904084-rs1686447	1.00E-02*	Yes
GGCX	Microsatellite-rs762684-rs6738645	1.78E-02	Yes
F7	rs3093229-rs3093233	2.42E-02	Yes
PDIA3	rs10163054-rs8040336	2.64E-02	Yes
PROZ	rs2273971-rs3024711	3.57E-02	Yes
F9	rs401597-rs392959	3.83E-02	Yes
NR1I2	rs2461818-rs7643645	3.93E-02	Yes
ORM1-2	rs1687390-rs3762055	4.93E-02	**No**

** Experiment-wise significance, significant p-value is 1.65E-04.
* Gene-wise significance, significant p-value depends on the effective number of test in each gene or gene cluster.
[a] Linkage disequilibrium with *CYP2C9**3 (rs1057910)
[b] Linkage disequilibrium with *CYP2C9**2 (rs1799853)

4.3 *VKORC1* AND *GGCX*

VKORC1 and GGCX are the two key enzymes in recycling vitamin K in our body. When *VKORC1*, the effect target for warfarin, was identified by Rost et al (Rost et al. 2004) and Li et al (Li et al. 2004), all publicly available SNPs (dbSNP 121) in this gene (including 5 kb up- and down-stream flanking regions) were selected. At that time, 29 *VKORC1* SNPs were examined; including the mutations reported by Rost et al. and 16 SNPs plus a microsatellite marker which was reported by Shikata et al (Shikata et al. 2004) on GGCX.

4.3.1 *VKORC1*

Out of the 29 tested SNPs, 20 passed the assay quality control with only five of them being polymorphic in the Uppsala cohort. This suggested that *VKROC1* does not tolerate much variation, but this may be also due to the lack of validation and allele frequency information in the SNP database. Sequencing efforts have been made by two independent studies to discover novel variants in *VKORC1* and flanking region (Geisen et al. 2005; Rieder et al. 2005). Geisen and colleagues sequenced 200 young healthy subjects (100 female and 100 male) on intragenic, 1.8 kb upstream and 1.5 kb downstream region and discovered four additional common SNPs with MAF ranging between 9.25% and 41.5%. Rieder and colleagues sequenced 186 warfarin-treated patients confirming 7 out of 8 SNPs reported by Geisen and colleagues. Meanwhile, Rieder et al also reported 3 additional SNPs in the 5' flanking region of *VKORC1*.which are in LD with the previously known common SNPs in this region.

Alleles of the five *VKORC1* SNPs co-vary significantly with warfarin dose according to the regression analysis. Four of them are extremely closely associated with dose, $P < 0.0001$. The fifth (rs11150606), which is located downstream of the gene and has a much lower MAF, shows a much less significant association, $P = 0.02$. All five SNPs are good predictors of warfarin maintenance dose, and explain 29 – 30% of inter-individual variability. Outside the above three SNPs (rs2359612 is the best predictor, Figure 4.1), the effect of rs7294 was evaluated in the regression model including rs2359612. The result showed that it explains a further 3% of the variance of warfarin maintenance dose.

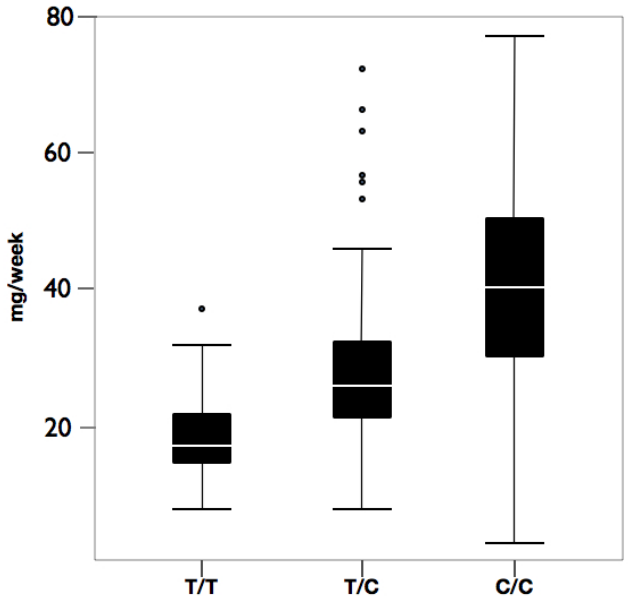

Figure 4.1. Mean weekly dose of rs2359612 genotype in 201 Swedish. The homozygous T/T is associated with low maintenance dose whereas C/C homozygotes need twice dose than T/T homozygotes.

Figure 4.2 shows the genomic structure, LD architecture, and common haplotypes of *VKORC1* and flanking regions. The arrow indicates the transcription from 5' to 3'. Four common SNPs have MAF of circa 40% and are located in the promoter region (rs9923231), the first intron (rs9934438), second intron (rs2359612) and 3' untranslated region (UTR)

(rs7294) of the gene. The fifth SNP (rs11150606) is located downstream of the gene and has a MAF of 4%. Inter SNP distances are 2.8, 1.1, 1.5 and 3.3 kb, respectively (Figure 4.2). We found that the four most common SNPs are in strong LD and give rise to three common haplotypes that are further subdivided into four by the more rare SNP rs11150606 (Figure 4.2). The first three SNPs are tightly linked, pairwise $r^2 = 1$.

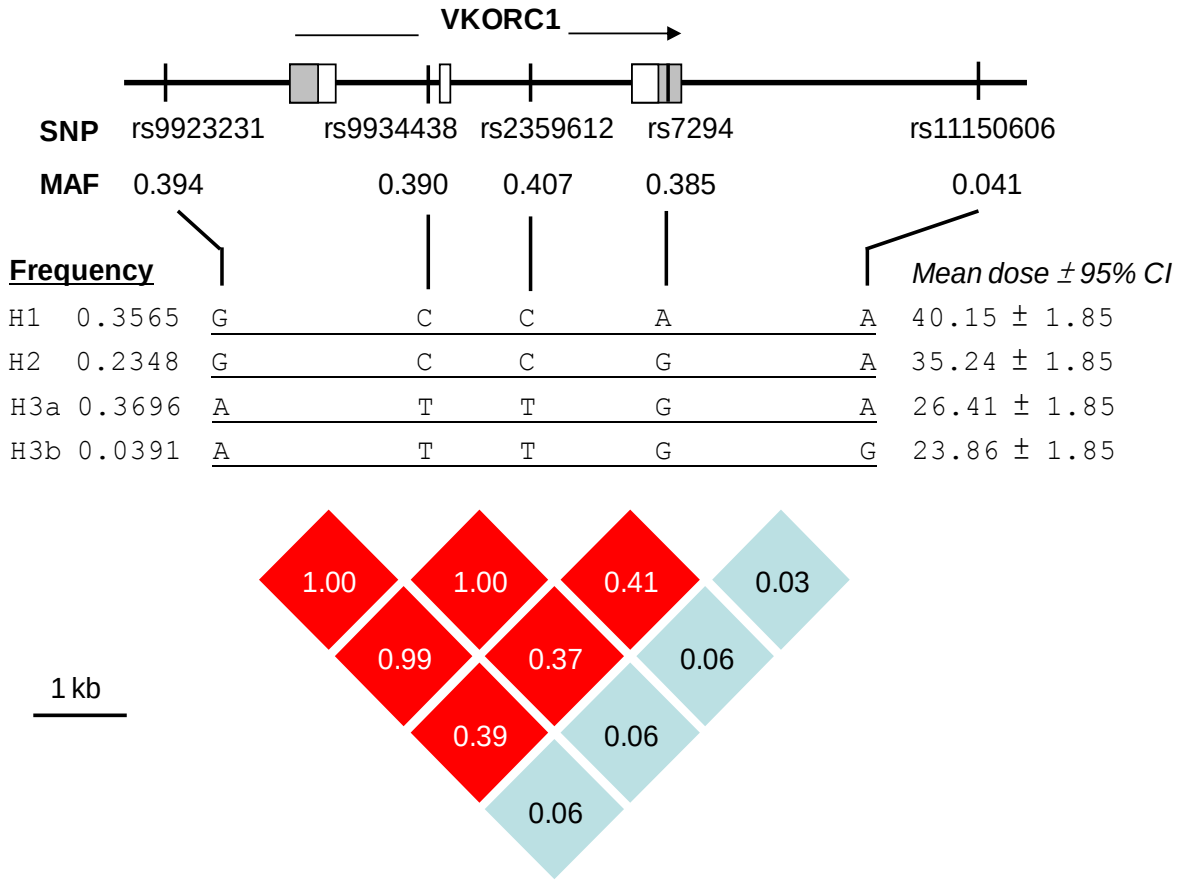

Figure 4.2. Genomic structure, haplotype, and linkage disequilibrium of *VKORC1*. Five SNPs were found to be polymorphic in Uppsala cohort and four of these are common (MAF ≥ 5%). The minor allele frequency (MAF) was noted under each SNP and haplotypes were listed (underline). The frequency of each haplotype and the corresponding mean weekly dose (and 95% confident interval) were calculated. The SNPs in 5' promoter region (rs9923231), intron 1 (rs9934438) and intron 2 (rs2359612) are tightly linked and only two haplotypes (GCC and ATT) were phased. Pairwise LD is represented in red (strong LD) and light blue (weak LD) diamonds with pairwise r^2 shown.

To characterise warfarin dose differences among *VKORC1* haplotypes, the means of each pair

of haplotypes were statistically compared (Figure 4.2 and Table 4.4) – confidence intervals were (CI) were also calculated using QTPhase component of Unphased software (Dudbridge 2003). Haplotypes that share alleles at the first three SNPs (rs9923231, rs9934438 and rs2359612) do not exhibit significant differences in warfarin dose; haplotype H1 G–C–C-A-A versus H2 G–C–C-G-A (P = 0.09), and A–T–T-G-A haplotypes H3a vs. H3b A–T–T-G-G (P = 0.54). However, every pair of haplotypes with different alleles at the first three SNPs (G–C–C vs. A–T–T) shows significant differences (P = $1.2x10^{9}$ to 0.0163). Figure 4.1 and Table 4.4 both illustrate that G–C–C haplotypes (first two rows) require significantly higher mean doses (35.24–40.15 mg) than the A–T–T haplotypes (last two rows, 23.86–26.41 mg). It is possible to subdivide further the high and low haplotypes but this distinction is associated with only 3% additional explanation in dose variation. The haplotypic results and the results from the univariate model discussed earlier indicated that the first three SNPs are the best predictors of warfarin dose, and that rs7294 and rs11150606 provide much less additional predictive information.

Table 4.4. Significance of pairwise haplotype comparison.

Haplotypes	H3b	H3a	H2	H1	Genotype
H1	0.00145198	1.26×10^{-9}	0.0913005	-	G-C-C-A-A
H2	**0.0163264**	**0.000117041**	-		**G-C-C-G-A**
H3a	0.540107	-			A-T-T-G-A
H3b	-				**A-T-T-G-G**

4.3.2 Fine mapping across the 500 kb LD block harbouring VKORC1 in Caucasians

The three SNPs rs2359612, rs9934438 and rs9923231 in *VKORC1* have nearly perfect LD in the Swedish population sample, and it is not possible with the type of analysis described above to dissect which one, if any of them, is causative. None of the reported coding SNPs

that were genotyped is polymorphic in these patients, nor has any other common coding variant been reported by re-sequencing studies totalling 772 chromosomes (Geisen et al. 2005; Rieder et al. 2005). In addition, none of the rare mutations reported by Rost et al (2004) was observed in this study. Hapmap results indicated that *VKORC1* is positioned at the right wing of an extended LD block spanning 285 kb in Caucasians (Figure 4.3). To corroborate the effect of *VKORC1*, all genes located in this large LD block were assessed. *STX4* is located 60 kb to the left of *VKORC1* (Figure 4.3) and encodes syntaxin 4 a peptide thought to be involved in haemostasis and the formation of a platelet plug (Reed et al. 1999). Therefore, it was necessary to investigate whether the observed association signal was due to variants in the *VKORC1* gene or not.

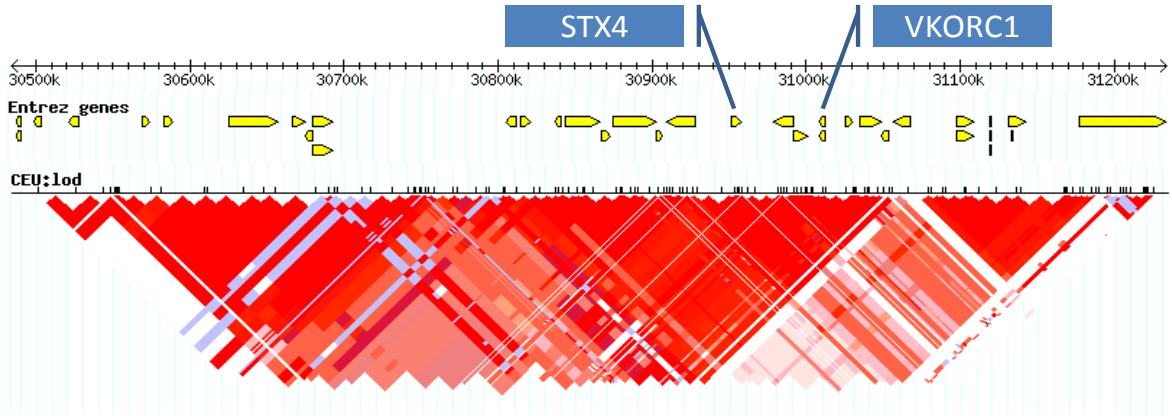

Figure 4.3. Genomic architecture of *VKORC1* region in Hapmap CEU panel (Caucasian).

To fine map the locus and elucidate *VKORC1* role in warfarin dose variation, a set of 13 additional polymorphic SNPs spanning the main and flanking LD blocks were selected from Hapmap (Figure 4.4). SNP genotypes were downloaded from HapMap website (phase I data) whereas LD, haplotype analyses and tag SNP selection were performed with the Haploview software (section 2.7.2). Manual SNP selection was performed if a tag SNP selected by Haploview failed in assay design (Sequenom platform).

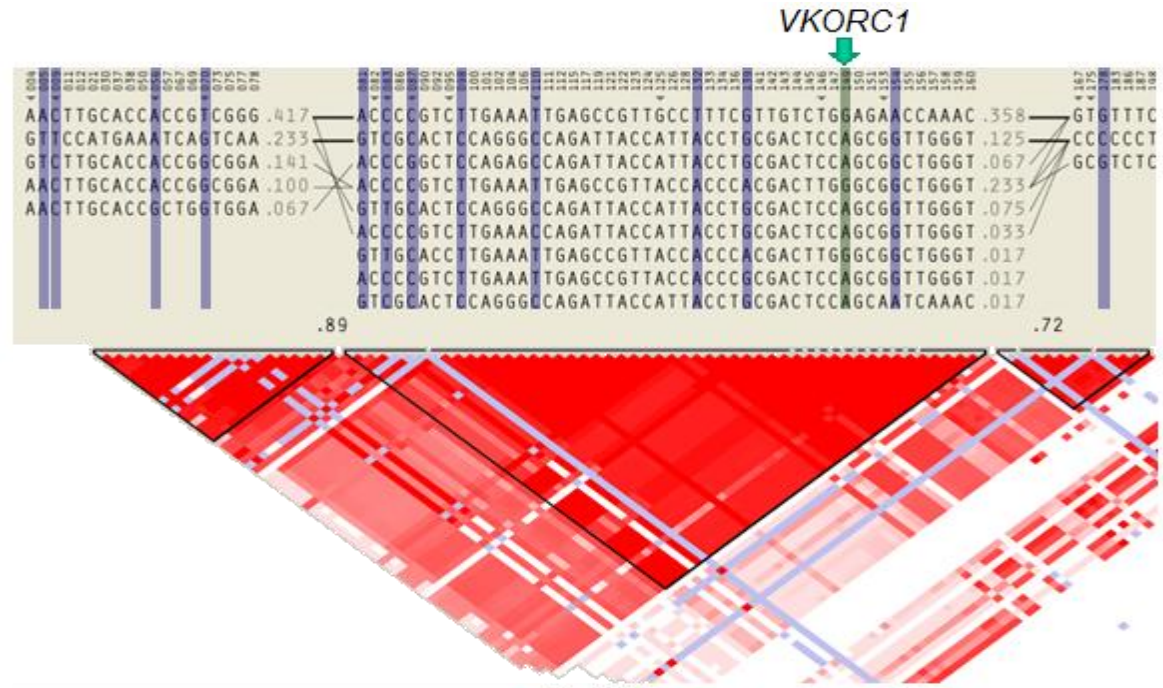

Figure 4.4. SNP selection for fine mapping *VKORC1* locus. SNP genotypes were downloaded from dedicated Hapmap website and haplotypes were phased with Haploview. SNP rs7294 located in *VKORC1* 3' UTR is shaded in green whilst selected tag SNPs are shaded in purple.

Ten SNPs in the main LD block which have either low or moderate r^2 with the four genotyped SNPs in *VKORC1* (Figure 4.5) and one SNP in each of the two flanking LD blocks were selected, giving a total span of ~558 kb. Finally, rs4889490 and rs4889599 which have moderate LD with rs2359612 and rs7294, respectively, were also selected to further refine the signal from VKORC1. Figure 4.5, panel B, shows that none of these SNPs gave higher association signal with warfarin dose compared to the original three linked SNPs which predicted circa 30% of the dose variance in our patients. The two SNPs rs4889490 and rs4889599 gave a signal slightly above the experiment-wise significance threshold due to their moderate LD with the four *VKORC1* SNPs. Further 2- or 3-marker haplotypic analysis indicated that no haplotype yielded association p-values more significant than the three linked SNPs. The fine mapping result confirmed the initial finding that common variants in *VKORC1* are the genetic determinants of warfarin dose.

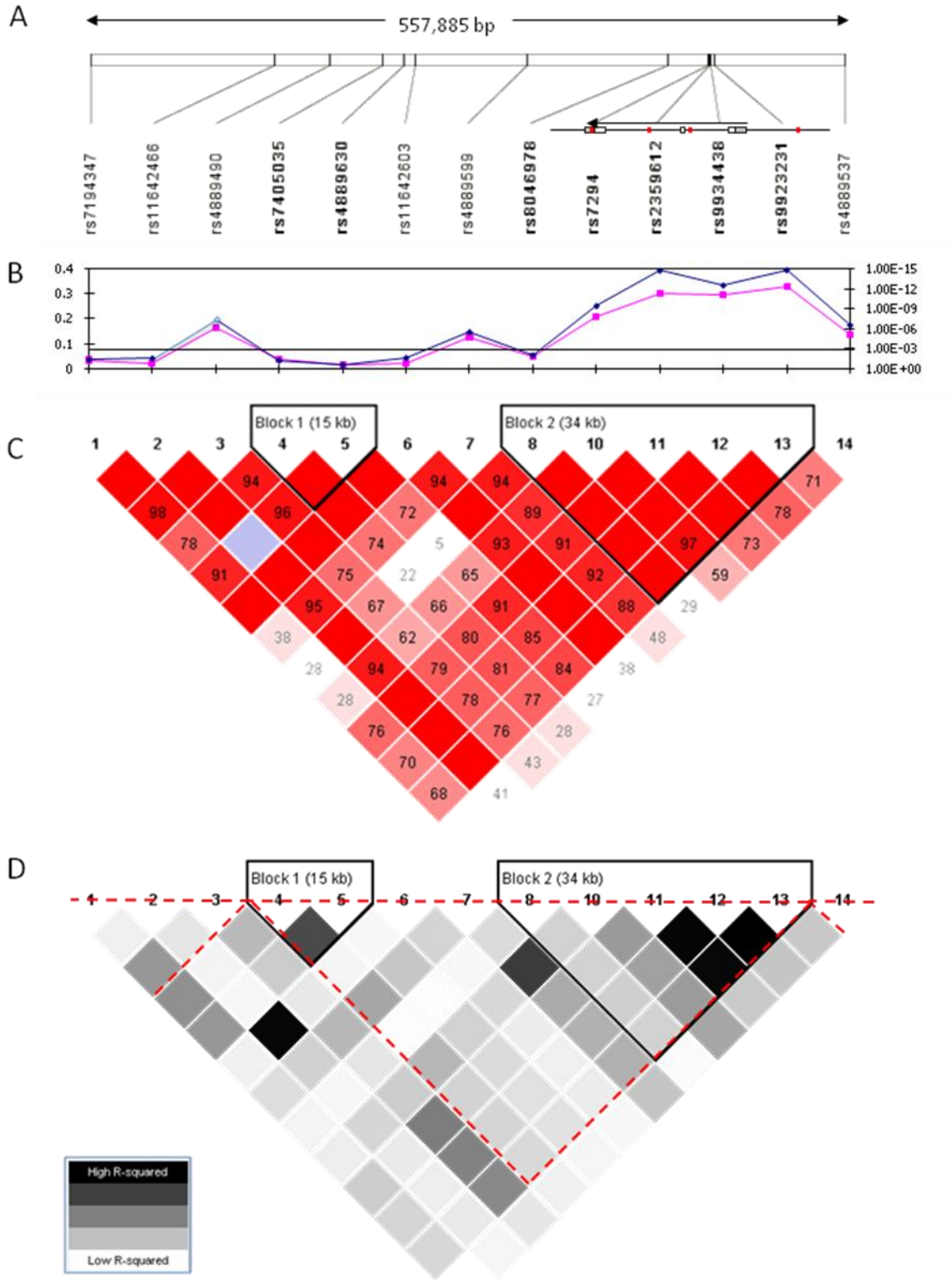

Figure 4.5. Association result of fine mapping *VKORC1* locus. (A) The SNPs flanking on LD blocks, including the four SNPs in *VKORC1*; (B) p-value (blue, right axis) and univariate R^2 (pink, left axis); (C) pairwise D prime and (D) pairwise r2.

4.3.3 The search of causative variants in *VKORC1*

To date there is no common coding sequence variant found in any of the published studies reporting the association of VKORC1 to warfarin dose requirement (D'Andrea et al. 2005; Rieder et al. 2005; Wadelius et al. 2005). Therefore, it is very likely that the causative variant is regulatory rather than one affecting the protein structure of VKORC1. So, does SNP rs9923231 have an effect on the transcribed messenger RNA (m-RNA) expression? This question was initially addressed by studies showing that rs9923231 is correlated with liver levels of mRNA (Rieder et al. 2005) and with activity in a reporter assay done in Hep G2 hepatoma cell lines (Yuan et al. 2005). These two studies both showed that the rs9923231 G allele is associated with higher mRNA expression (Figure 4.6). However, Bodin et al (2005) failed to confirm this reporter assay with the same experimental setup described in Yuan et al.

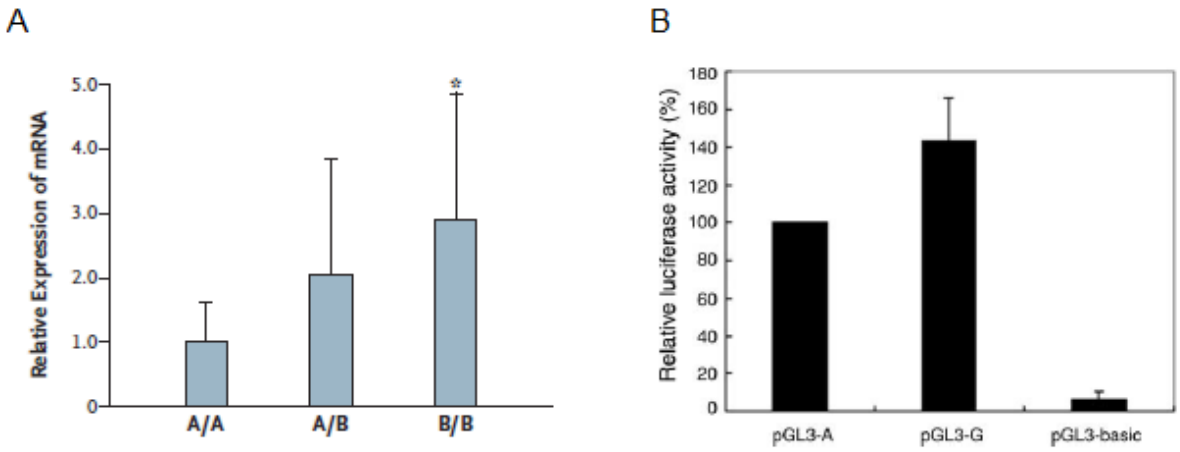

Figure 4.6. *VKORC1* mRNA expression assays. (A) Correlation between *VKORC1* haplotype groups and mRNA expression and (B) Luciferase activity assay with *VKORC1* promoter region. (A) the relative mRNA expression of *VKORC1* in liver was measured accordingly haplotypes A and B. The haplotype A carries minor A allele of rs9923231 whereas haplotype B carries the major G allele. (B) pGL3 luciferase reporter containing either the A (pGL3-A) or the G allele (pGL3-G) of rs9923231 whereas pGL3-basic is negative control without inserted sequence. Figure A is reproduced from N Engl J Med 2005 352:2285-2293 (Rieder et al. 2005) whilst Figure B is reproduced from Hum Mol Gen, 2005, 14(13):1745-1751 (Yuan et al. 2005).

Although the above preliminary results suggested that rs9923231 might be functional in regulating the expression of VKORC1, a further effort to confirm this observation is necessary due to the contradictory results reported by Bodin et al. A collection of liver biopsy samples from 25 individuals were provided courtesy of Dr Ana Alfirevic (Liverpool). The samples were used to test the mRNA expression in association with *VKORC1* genotypes. The RNAs of each liver biopsy were extracted with TRIZOL reagent and thereafter tested with quantitative PCR (Applied Biosystems). However, due to the age of these liver biopsy samples, the extracted RNAs were degraded and it was not possible to obtain reliable results (Figure 4.7). A second attempt was also made by using the Illumina DASL assay which was also unsuccessful (Dr Matthew Forrest).

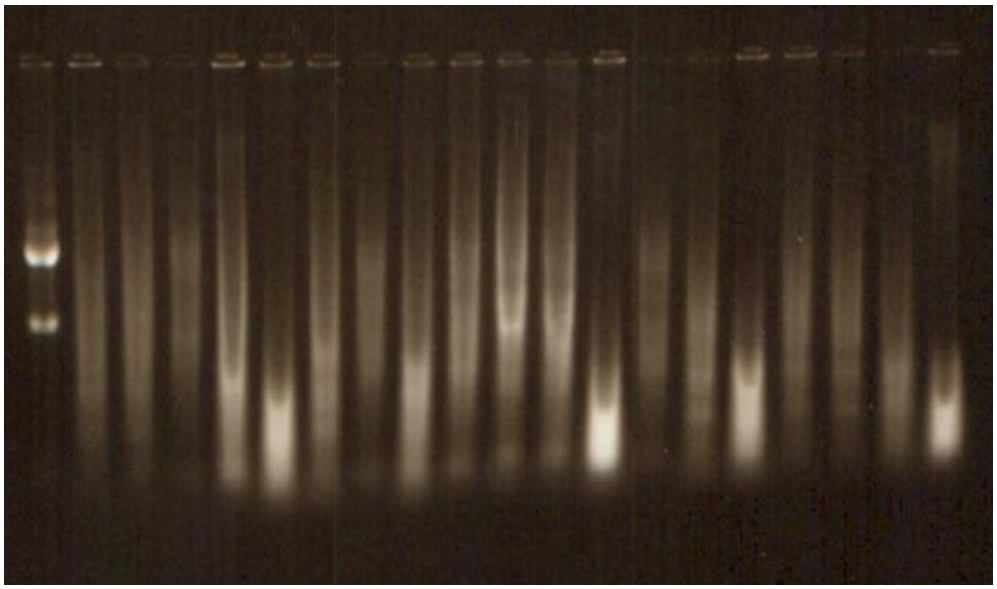

Figure 4.7. Electrophoresis of RNA samples prepared from the liver biopsies. The first lane indicates 18S (lower band) and 28S (upper band) standards.

Comparative genomics has demonstrated its powerfulness in predicting functional important segments, such as coding regions and gene regulatory elements resulting from purifying selection in evolution (Pennacchio and Rubin 2001). Numerous tools have been developed

for comparative sequence analysis such as MultiPIP (Schwartz et al. 2003) and zPicture (Ovcharenko et al. 2004). The corresponding sequences from chimpanzee, mouse, rat, dog, and chicken, corresponding to the human *VKORC1* locus, were downloaded from the Ensembl database and multiple sequence alignment was performed with zPicture (Figure 4.8).

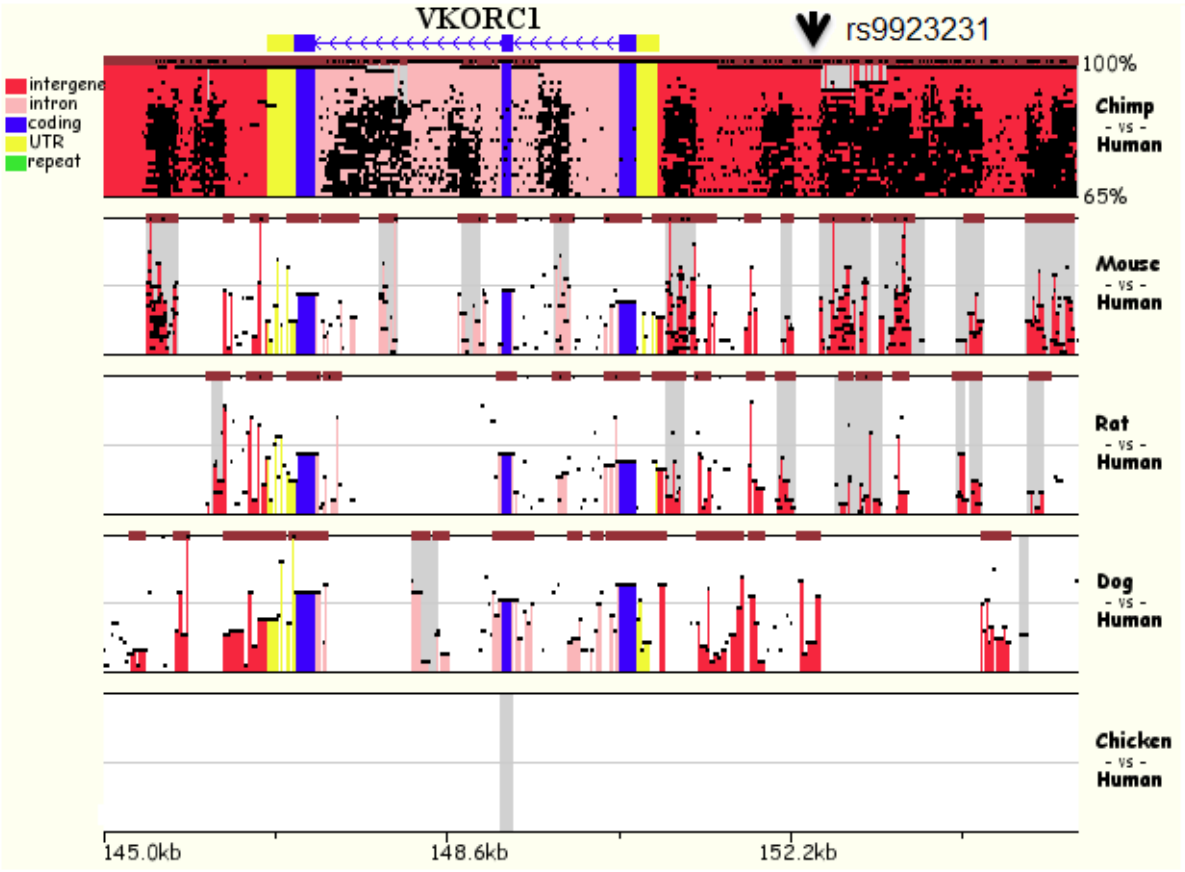

Figure 4.8. Genomic sequence alignment in *VKORC1* locus of human, chimpanzee, mouse, rat, dog and chicken. The black arrow indicates the location of potential functional SNP rs9923231.

Figure 4.8 shows the exon-intron structure of *VKORC1* at the top, whereas the genomic location of the putative promoter SNP rs9923231 is indicated with a black arrow. The three exons are conserved in all species except chicken whereas no significantly conserved region is found outside the genetic region; including the region harbouring rs9923231. This result

suggests that a comprehensive scanning for regulatory elements will be a good alternative approach to elucidate the expression correlation with *VKORC1* genotype.

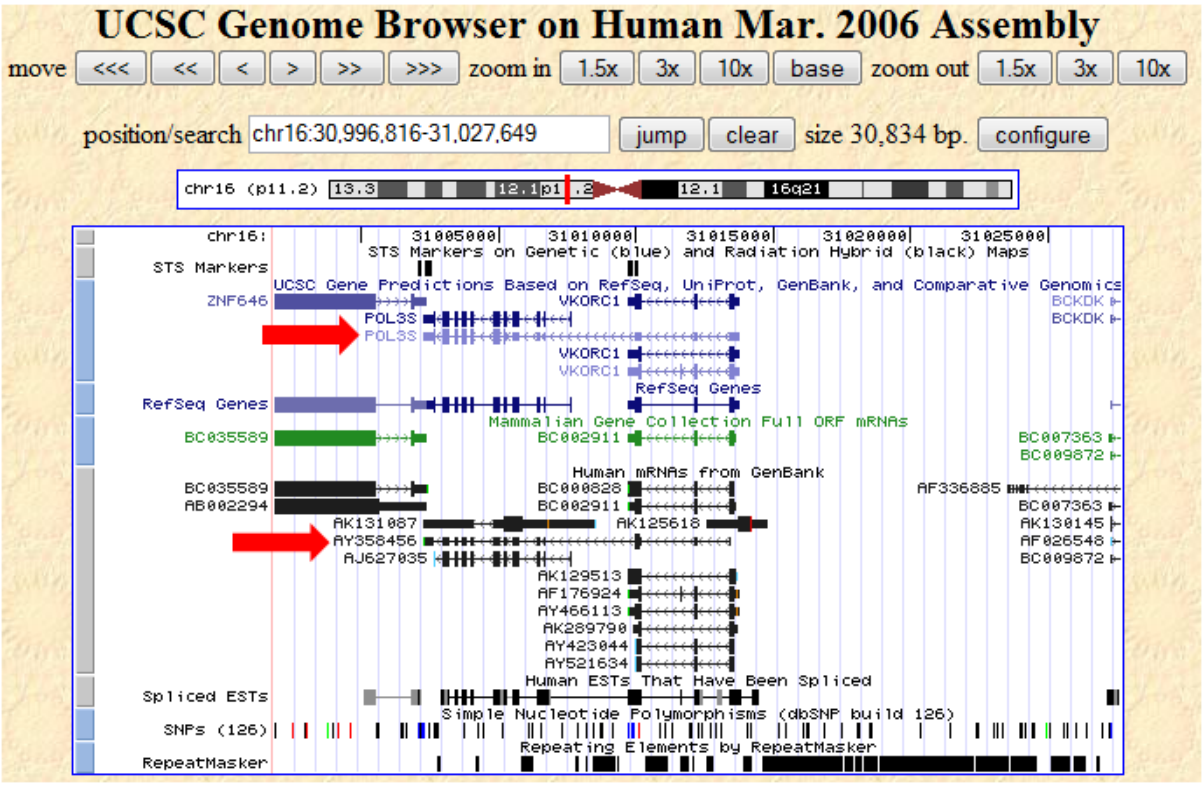

Figure 4.9. Alternative splicing variants of *VKORC1*. The red arrow indicates the evidence of a joint transcript of *VKORC1* and *POL3S* from a full length mRNA AY358456.

A few alternative splice variants have been reported in public databases based on the information of sequencing of full length cDNA libraries and expressed sequence tags (ESTs). Although none of these variants have been confirmed, interestingly, a long transcript which includes all three exons of *VKORC1* and all but the first exon of *POL3S* (polyserase) is supported by a full length mRNA transcript AY358456. The biological relevance and coding potential of such a transcript is unknown. It is worth mentioning that experimental work undertaken in our laboratory confirmed the presence of such transcripts in cDNA libraries from liver and lung tissue.

4.3.4 GGCX

GGCX acts in carboxylating vitamin K dependent proteins including F2, F7, F9, F10, protein C, protein S, and protein Z. For *GGCX*, we obtained results for 9 SNPs that passed study criteria and one microsatellite (Table 3.14). Figure 4.10 shows the LD architecture of the *GGCX* locus based on the nine common polymorphic SNPs, MAF above 30%. They are located in intron one (rs7568458), intron two (rs12714145), intron five (rs6738645), intron six (rs762684), exon eight (rs699664), exon nine (rs2592551), intron 14 (rs2028898) and in the 3' flanking region (rs6547621 and rs7605975). The exon eight SNP rs699664 leads to an arginine to glutamine change in codon 325, whilst rs2592551 in exon nine is a synonymous SNP. Inter-SNP distances are 0.8, 4.2, 1.1, 1.5, 0.4, 2.9, 2.6 and 2.1 kb, respectively. All nine SNPs are within a region of strong LD and define five common haplotypes (Figure 4.10).

In contrast to *VKORC1*, only one of the *GGCX* SNPs that passed study criteria reaches nominal statistical significance, rs12714145 (intron 2), P = 0.0360 (Figure 4.11). GGCX SNPs rs762684 (intron 6) and rs2592551 (exon 9) also show a tendency towards association with warfarin dose (P = 0.0613 and 0.0870). The mean warfarin dose associated with each haplotype was also calculated. A global test for statistical difference among the haplotype means revealed that there is no dose association among these haplotypes (P = 0.757).

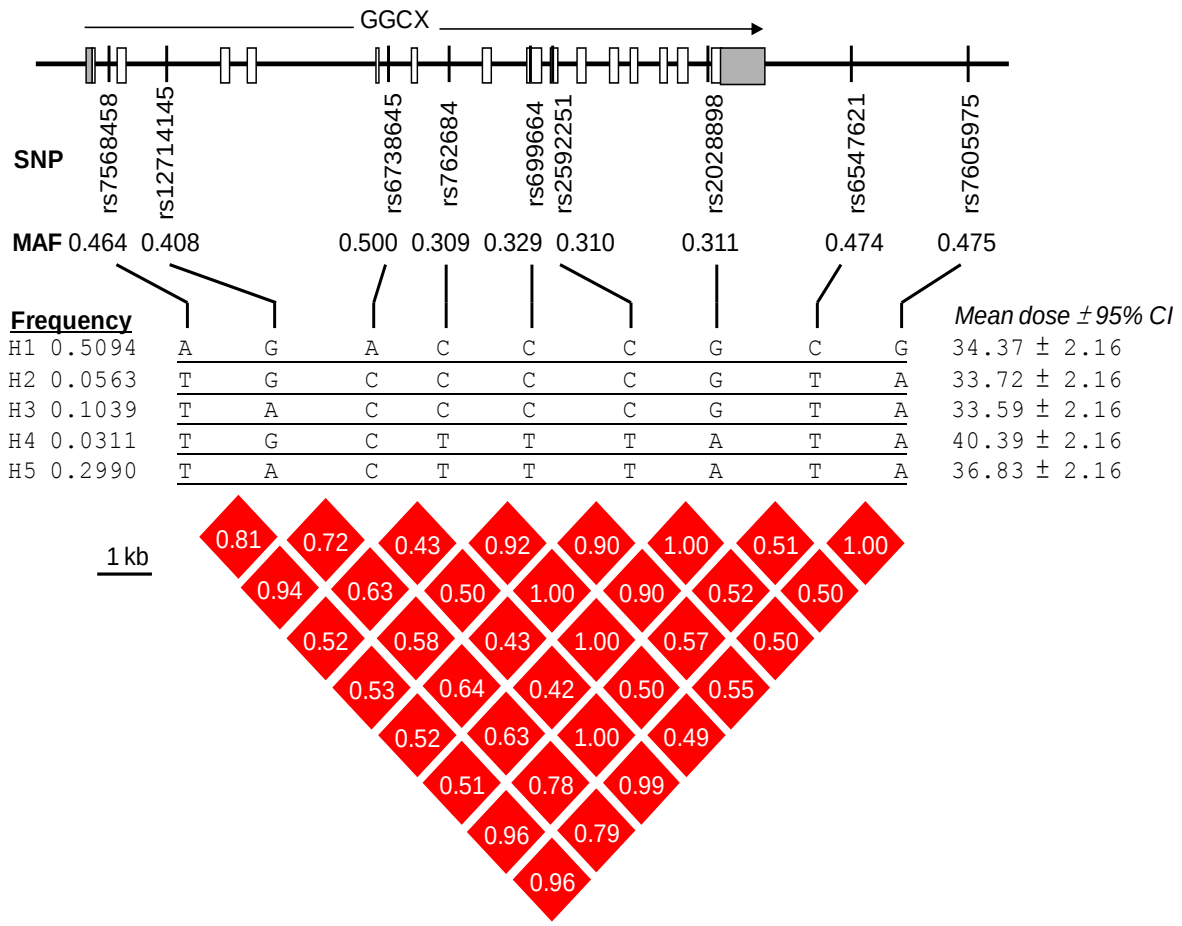

Figure 4.10. Genomic structure, haplotype, and linkage disequilibrium of *GGCX*. Nine SNPs were found polymorphic and common in Uppsala cohort. The MAF was noted under each SNP and ranges from 0.310 to 0.5. Haplotypes were listed with underline with annotation of frequency and mean dose (weekly) and 95% confident interval (CI). All SNPs are in strong linkage disequilibrium. Pairwise LD is represented in red (strong LD) and light blue (weak LD) diamonds with pairwise r^2 on it.

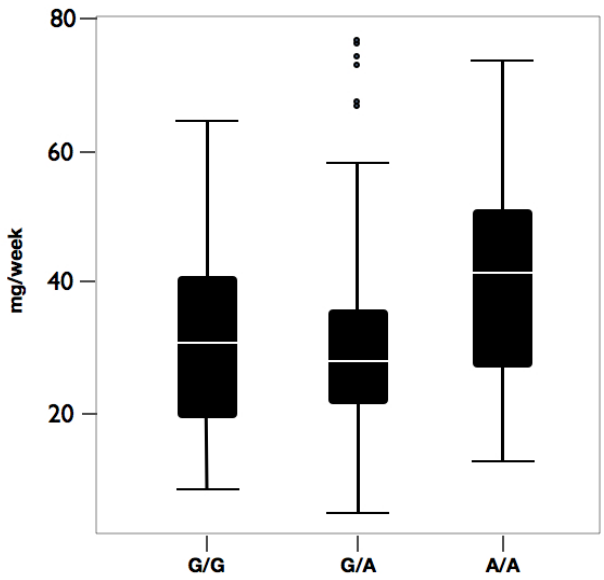

Figure 4.11. Mean weekly dose of rs12714145 genotype in 201 Swedish.

At the time of the above SNP analysis, Shikata et al (Shikata et al. 2004) described a microsatellite marker in intron 6 of the *GGCX* gene that was associated with warfarin dose. Three alleles with 10, 11, and 13 (CAA) repeats were detected in the 45 warfarin-treated Japanese patients. Three individuals which were heterozygous for the 13-repeats allele (10/13 or 11/13) required higher maintenance dose than patients carrying only alleles with 10 and/or 11 repeats. When this marker was typed in the Uppsala cohort, a wider range of alleles than in the Japanese cohort: was detected. Alleles with 10, 11, 13, 14, 15, or 16 repeats were found, with the 10-repeat allele being the most common. In analogy with the Shikata study, patients were divided into groups according to genotype: (1) 10/10 repeats, (2) 10/11 or 11/11 repeats, and (3) 10/13 or 11/13 repeats. In addition, a fourth group of patients has a greater number of (CAA) repeats, that is, homozygous for 13 or heterozygous for 14, 15, or 16 repeats (Figure 4.12).

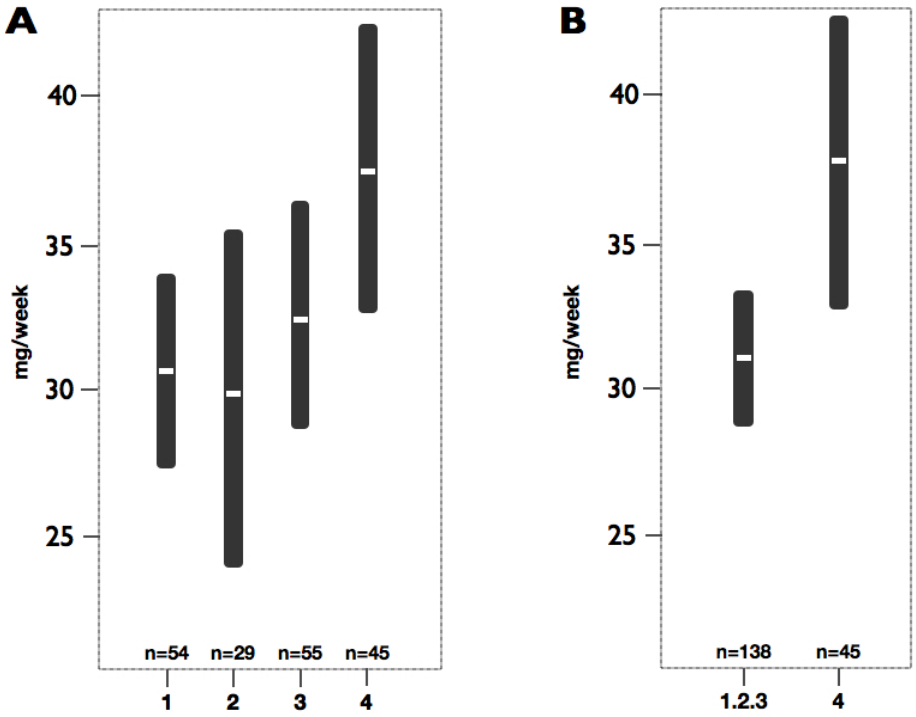

Figure 4.12. Individuals are divided into 4 groups. According to *GGCX* $(CAA)_n$ microsatellite genotype. Group 1: 10/10 repeats; group 2: 10-11/11 repeats; group 3: 10-11/13 repeats; group 4: 13/13 or x/14-16 repeats, x could be any repeat genotypes. (A) Mean weekly dose with 95% CI for four individual group. (B) Combined group 1-3 vs. group 4.

As in Shikata et al, warfarin dose requirement in the Uppsala cohort tends to increase with the number of microsatellite repeats, although the effect is only apparent in patients with higher numbers of repeats (group 4; Figure 4.12A). A combined of group 1, 2 and 3 against the fourth group shows nominally significant association (P = 0.011, Figure 4.12B). A *GGCX* polymorphism, rs12714145, has previously been observed in intron 2 that was associated with an increase in warfarin dose requirement (P = 0 .036). In chapter V, this finding wil be revisited in light of results in the much larger WARG study (see section 5.3.3).

4.4 CYP2C CLUSTER

In the previous chapter the eight P450 genes interrogated as part of this study were discussed (section 3.3). Four of them *CYP2C18, CYP2C19, CYP2C9* and *CYP2C8* constitute the CYP2C cluster on chromosome 10 (Figure 4.13). In the HapMap CEU panel, there is an extended region of strong LD harbouring these four genes. As shown in Figure 4.13 there is some LD granularity in this region with one large LD block harbouring *CYP2C18*, *CYP2C19* and the exonic part of *CYP2C9*, which includes all known functional *CYP2C9* variants, whereas a smaller LD block harbours *CYP2C8* and the flanking region of *CYP2C9* (Figure 4.13).

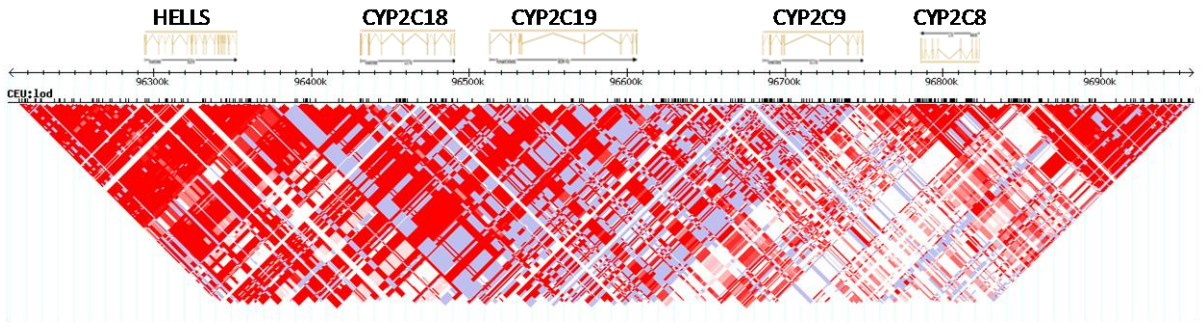

Figure 4.13. Genomic architecture of CYP2C gene cluster on chromosome 10.

Prior to this thesis, the *CYP2C9* gene had already been reported to be associated with warfarin dose (Aithal et al. 1999; Higashi et al. 2002). As expected, several CYP2C SNPs were found to be associated with warfarin dose. These were significant even after correction for multiple testing (Tables 4.2, 4.3).

The association data presented suggested that the CYP2C gene cluster on chromosome 10 was the second most strongly associated region after *VKORC1* (Tables 4.2 and 4.3). *CYP2C8,*

CYP2C9, and *CYP2C19* have been intensively studied and numerous genetic variants which alter amino acid composition have been reported (Human Cytochrome P450 Allele Nomenclature Committee, http://www.cypalleles.ki.se/).

To date *CYP2C9* has 30 reported alleles which may either increase or decrease its enzymatic activity in metabolising drugs. Among these variants, the *2 and *3 alleles have been reported to associate with warfarin dose in different populations (Solus et al. 2004). The functional *CYP2C9**2 polymorphism (rs1799853, R144C) confers a moderate decrease in the metabolism of S-warfarin (Rettie et al. 1994) and was not significant in our univariate analysis in the Uppsala cohort. The *CYP2C9**3 (rs1057910, I359L) severely impairs the efficacy to the hydroxylation of S-warfarin (Sullivan-Klose et al. 1996) and was the most strongly associated SNP in this region in our study (Table 4.2).

In this 400 kb region of high LD, 55 SNPs within, or flanking, the CYP2C cluster passed study criteria and were included for further analysis: 17 in *CYP2C9*, 10 in *CYP2C19,* 14 in *CYP2C18* and 12 in *CYP2C8* (Table 3.7 and Figure 4.14). The genomic architecture in the Swedish sample is similar to that reported by Hapmap, i.e. *CYP2C18, CYP2C19* and *CYP2C9* are in one large block whereas *CYP2C8* is located in separate smaller block. The LD blocks illustrated in Figure 4.14 are according to the block definition described by Gabriel and colleagues (Gabriel et al. 2002). However, an extended LD block can be clearly seen using less stringent thresholds (Ahmadi et al. 2005).

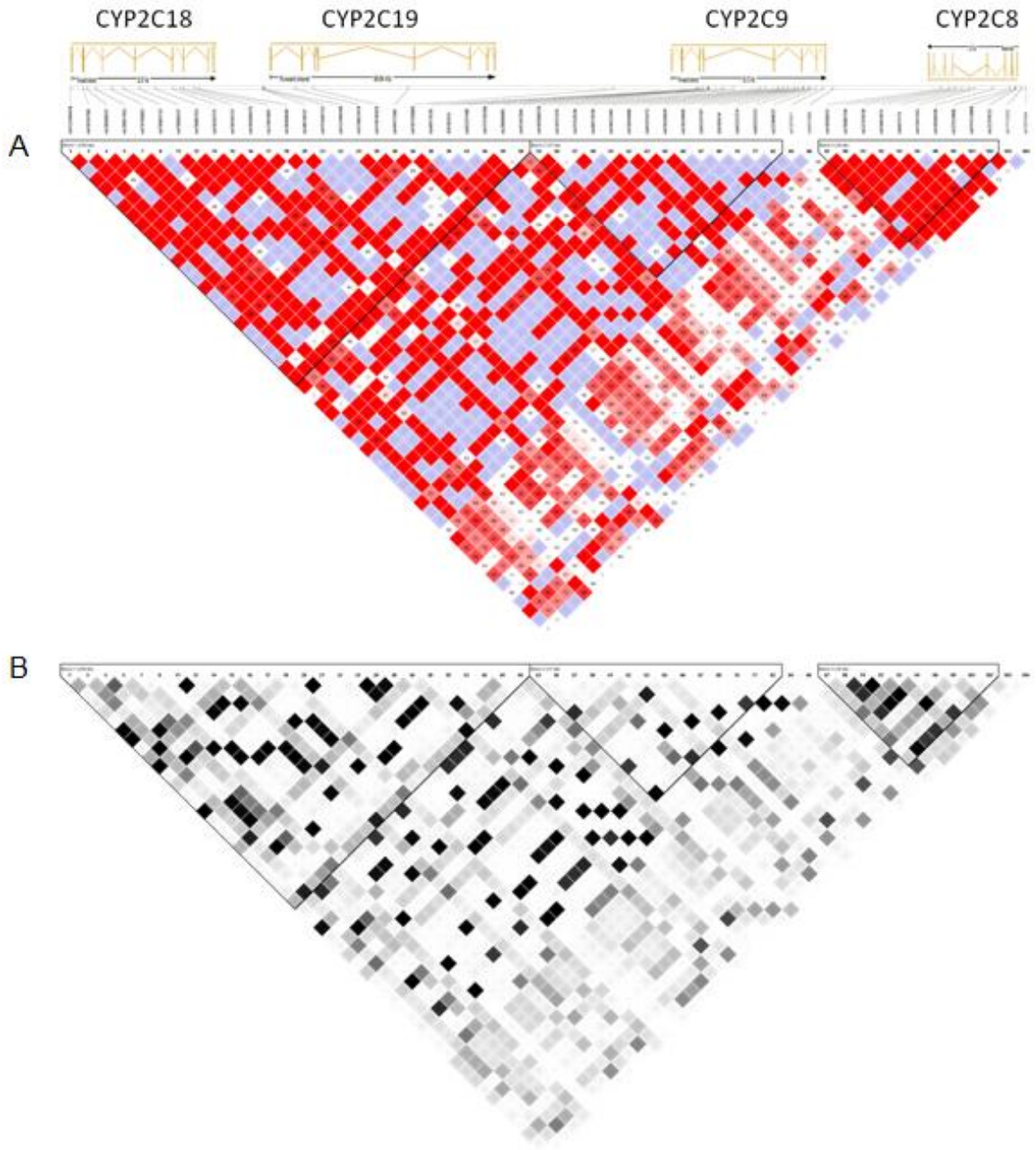

Figure 4.14. Genomic architecture of CYP2C gene cluster in 201 Swedish for (A) Pairwise D prime and (B) Pairwise r^2 calculation. The block is defined with definition described by Gabriel and colleagues (Gabriel et al. 2002).

In univariate analysis, nine SNPs showed significant association. To further explore the dose association of these nine SNPs, two multiple regression models were applied. Both models contained *VKORC1* (rs2359612) and significant non-genetic predictors of warfarin dose (Age, body weight and interacting drugs), whereas one contained *CYP2C9**2 but not *3 and vice

versa.

Each of the 53 SNPs in the CYP2C region was evaluated against the two multiple regression models. The result indicated that all significant results were fully explained by LD with either *CYP2C9**2 or *3, except for rs4917639 in *CYP2C9* (Table 4.5). SNP rs4917639 demonstrated a significant P-value in both the *2 model ($P < 1.33x10^{9}$) and the *3 model ($P < 3.56x10^{3}$). However, haplotype analysis showed that the minor allele of rs4917639 was in perfect LD ($r^2 = 1$) with a composite minor allele formed by aggregating *CYP2C9**2 and *3 into a single allele; although *2 and *3 are rarely carried on the same haplotype. This result suggests that the rs4917639 mutation occurred first, and the *2 and *3 alleles arose independently on the same parent allele. The strong association between rs4917639 and *2/*3 could perhaps be due to positive selection; if rs4917639 lessened the deleterious effect of impaired CYP2C9 metabolism caused by *2 and *3.

Table 4.5. Nominally significant result in multivariate regression model for SNPs in CYP2C gene cluster.

Gene	SNP	R^2 with *2 model	p-value	R^2 with *3 model	p-value
CYP2C8	rs11572080	0.001	0.502	0.017	0.005
CYP2C9	rs9332108	0.109	1.57E-10	0	0.999
CYP2C9	rs1057910 (*3)	0.109	1.57E-10	-	-
CYP2C9	rs1057911	0.114	2.97E-10	0	0.999
CYP2C9	rs9325473	0.112	5.45E-10	0	0.999
CYP2C9	rs4917639	0.1	1.33E-09	0.025	3.56E-03
CYP2C9	rs9332214	0.098	1.50E-09	0	0.999
CYP2C9	rs2860905	0.048	1.86E-04	0.009	0.153
CYP2C9	rs4917636	0.004	0.213	0.026	3.63E-03
CYP2C9	rs4607998	0.004	0.252	0.026	2.85E-03
CYP2C9	rs1799853 (*2)	-	-	0.024	4.00E-03
CYP2C9	rs1934966	0	0.999	0.015	8.72E-03
CYP2C9	rs9332222	0	0.999	0.025	3.86E-03
CYP2C18	rs7896133	0.063	5.17E-07	0	0.999
CYP2C18	rs2901783	0.02	0.029	0.004	0.471
CYP2C19	rs3814637	0.098	3.76E-09	0	0.896
CYP2C19	rs17882687	0.047	5.19E-05	0	0.828

In conclusion, the poor metabolising effect of *CYP2C9**3 alleles in the Swedish sample was confirmed. Compared to *CYP2C9**3, *CYP2C9**2 showed a minor effect in warfarin maintenance dose in the Swedish cohort. In the Uppsala cohort, no independent effect is identified except *2 and *3 alleles and rs4917639 which is linked to both alleles.

4.5 OTHER NOMINALLY ASSOCIATED GENES

Except *CYP2C9* and *VKORC1* which directly interact with warfarin, *PROC* emerged as the most likely factor which may influence the dose variation. In this study, 13 SNPs in *PROC* passed study criteria (Table 3.21.). In univariate analysis, four out of the 13 SNPs tested were significantly associated with dose (Table 4.2). The associated SNPs were located in the 5' regulatory region (rs1799809 and rs2069901), in intron 2 (rs2069910), and in intron 3 (rs2069919). Except for rs2069910, all other SNPs reached gene-wise significance. Haplotype analysis of the 13 SNPs in *PROC* did not provide a stronger signal compared to single marker analysis, although the haplotype derived from rs2069919, rs2069921 and rs973760 reached gene-wise significance (P = 0.00136). In the model with *VKORC1*, *CYP2C9*, body weight and interacting drug, the two SNPs in the promoter region of *PROC* explained 7–9% of the variance in warfarin dose. Previous studies had suggested that homozygotes for the G allele of rs1799809, which is located in the promoter region of *PROC,* have slightly decreased enzymatic activity and reduced levels of PROC concentration (Aiach et al. 1999; Spek et al. 1995). Furthermore, in 2003, Watala et al reported a correlation between the enzymatic activity of PROC and the coagulation rate both in patients treated with oral anticoagulant and in healthy volunteers (Watala et al. 2003). Our results suggested that patients homozygous for the G allele required a lower dose, which is in agreement with the above finding in biological function of protein C. The PROC association reached only gene-wise significance and in a small sample of 201 patients.

Other than *PROC, APOE* was the only other gene that reached gene-wise significance in the univariate analysis for association with warfarin dose. The two *APOE* SNPs (rs429358 and rs7412) that discriminate the widely described *E2, *E3 and *E4 allelic system showed that

patients who carry the common allele *E4 or the rarer *E2 require higher warfarin doses than those with *E3 (Table 4.2). This finding is contradictive to a UK study (Sconce et al. 2006). A Dutch study analysing patients taking the anticoagulant drug phenprocoumon showed that the *E4 is associated with higher maintenance dose, but the *E4 allele carriers required lower maintenance doses of acenocoumarol (Visser et al. 2005). This discrepant result suggests that the association between anticoagulant dose and APOE may result from sampling bias, and a larger cohort treated with phenprocoumon, acenocoumarol and warfarin will help to refine this association (Wadelius et al. 2007).

Besides the genes that have been described above, eight additional genes were nominally associated with warfarin dose but did not pass either the experiment-wise or gene-wise threshold after correction for multiple testing: *EPHX1*, *CALU*, and *ORM1-2* in single marker analysis and *F7*, *PROZ*, *F9* and *NR1I2* in haplotype analysis (Tables 4.2 and 4.3).

In single marker analysis, rs1051740 in *EPHX1* reported in association with high maintenance dose (Loebstein et al. 2005) is not replicated, but, instead, another SNP rs4653436 is nominally significant. A functional variant in *CALU* reported to potentially increase warfarin dose is also not replicated. However, four other SNPs in *CALU* show nominal significance. SNPs in *PDIA2* and *ORM1-2* showed a nominally significant p-value and marginal association with dose. To date, there has not been any report with *ORM1* and *ORM2* in association with warfarin dose.

Haplotypes in *PDIA6* showed gene-wise significance in association with dose whereas in *PDIA3* it is nominally significant (Table 4.3). Although PDI has been demonstrated to be involved in providing electrons to the thioredoxin-like CXXC centre in VKORC1 and in

initiating tissue factor pathway for coagulation, precisely which PDI member is involved in these interaction remains unclear (Ahamed et al. 2006; Versteeg and Ruf 2007; Wajih et al. 2007). The haplotypes in *F7* and *PROZ* showing nominal significance are both comprising SNPs located in the upstream flanking region (Table 4.3), and, interestingly, *F7* and *PROZ* reside back to back on chromosome 13 with *F10*. Our finding of *F7* upstream haplotype is in accordance with previous studies which have reported variations in the *F7* promoter region to be associated with warfarin sensitivity (Aquilante et al. 2006; D'Ambrosio et al. 2004; Shikata et al. 2004). Shikata and colleagues also reported SNPs rs5896 in *F2* and rs5960 in *F10* to be associated with warfarin dose (Shikata et al. 2004), but these observations did not replicate in our Uppsala study. In *F7*, the functional variant R413Q, (rs6046) which was reported to be associated with decreased plasma levels of F7 (Arbini et al. 1994), did not replicate either in our study. A rare variant (A-10V/T) in *F9* which was previously reported to be associated with decreased F9 enzymatic activity, was not polymorphic in the Uppsala sample. However, a haplotype spanning this exon has shown a marginal association. Finally, the haplotype on *NR1I2* is located near the one of the two recombination hot spots and may harbour a novel mutation which influences warfarin dose. To date, no variant in *NR1I2* has been reported in the literature to be associated with warfarin dose and replication in an independent study will validate this finding.

Although some of the variants described above appear to be marginally significant and explain the variability of dose variation in the Uppsala cohort, they should be treated as provisional at this stage pending replication in a larger cohort of the same ethnic origin. This work is described in Chapter V.

4.6 PREDICTIVE MODEL FOR WARFARIN DOSE

Besides the genes showing significant association with warfarin to various degrees, non-genetic factors also contribute to dose variability and age, body weight, and drug interaction (concomitant medication) can better explain the dose variation (Table 4.6). To develop the best multiple regression model, these factors were taken into account with the genetic factors showing at least nominal significance.

Table 4.6. Association test of non-genetic factors.

Variables	P-value	r^2
Age	0.0029	0.092
Bodyweight	0.0075	0.057
Interaction	0.0878	0.036
Gender	0.0314	0.023
Indication	0.0819	0.015
PT INR	0.1272	0.012

Niclas Eriksson (Uppsala) developed a multiple model in which both non-genetic and genetic factors were assessed. For non-genetic factors, age, body weight, interacting drugs, and indication were retained while gender and PT INR value were excluded in the multiple model. In Figure 4.15, the contribution of genetic factors into different models based on the strength of the association signal is shown. Together with non-genetic factors, the two genes showing experiment-wise significance, *VKORC1* and *CYP2C9*, account for 56.0% of the total inter-individual variation in warfarin response. If *PROC* is included in the model with *VKORC1* and *CYP2C9*, 62% dose variation could be explained for the variance in dose (Figure 4.15).

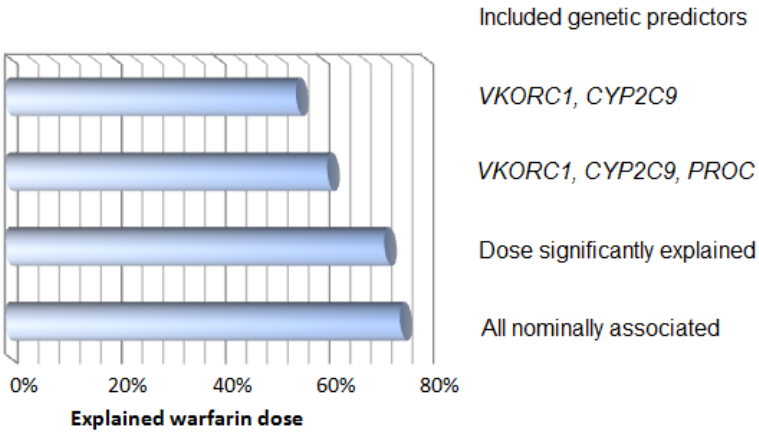

Figure 4.15. Multiple models of dose explained by genetic and non-genetic factors. Different models comprised different combination of predictors were tested to explain the inter-individual dose variation.

Finally, to explore the full potential of all findings from the Uppsala study in describing dose variation, all genes showing nominally significant association including *VKORC1*, *CYP2C9*, *CYP2C19*, *CYP2C18*, *PROC*, *APOE*, *EPHX1*, *CALU*, *GGCX*, *ORM1-2* and non-genetic factors were recruited in the multiple model and a surprising 76% of the inter-individual variation of warfarin dose could be explained (Figure 4.15). However, to refine this model, genetic factors with p-values above 0.2 were removed in a stepwise fashion, together with low-explanatory value (R^2) from the model. A final model which includes *VKORC1*, *CYP2C9**2 and *3, *PROC*, *EPHX1*, *GGCX*, *ORM1-2*, age, body weight, and drug interactions explains 73% of the dose variation (Figure 4.15; Table 4.7).

Table 4.7. Predictors in final multiple model of warfarin dose variation.

Predictor	SNP	p-value	Univariate R^2
VKORC1	rs9923231	<.0001	0.317
CYP2C9	rs1799853 (*2) + rs1057910 (*3)	<.0001	0.159
Age		0.0029	0.092
PROC	rs2069919	0.0416	0.09
Bodyweight		0.0075	0.057
EPHX1	rs4653436	0.1016	0.048
Drug interaction		0.0878	0.036
GGCX	rs12714145	0.026	0.034
ORM1-2	rs1687390	0.0571	0.026

Although p-values of *EPHX1*, *ORM1-2* and drug interaction do not reach even nominal significance, these predictors could explain a substantial variation of dose. Since this final model is evaluated based on a small study sample, validation of these effect would be necessary to confirm the influences.

4.7 CONCLUSIONS

A comprehensive selection of 35 candidate genes was investigated in association with inter-individual dose variation. The result suggests that *VKORC1* and *CYP2C9**2 and *3 are significantly associated with the effect of dose variation. This finding in our Swedish sample was also replicated in different ethnic populations in other studies (D'Andrea et al. 2005; Rieder et al. 2005; Yuan et al. 2005). The outcome of CYP2C9*2 and *3 alleles is well studied and both alleles are known to impair CYP2C9 enzymatic activity. However, the biological explanation of VKORC1 action is still unclear although the two *CYP2C9* alleles explain well the dose variation and preliminary evidence suggests a variation in mRNA expression (Rieder et al. 2005; Yuan et al. 2005). Other possibilities include alternative splicing or remote regulatory elements which require further investigation including functional studies.

A conservative multiple model including age, body weight, drug interaction, genotypes of VKORC1 and CYP2C9*2 and *3 is able to explain 56% whereas additional predictor of PROC contributes another 6% with a total of 62%. This model could be used to develop a dosing algorithm and tested in prospective studies. We found additional, marginal associations which may be population specific. At this point, these findings were tested for replication in a larger cohort; which will be covered in the next chapter.