CHAPTER III

SELECTION OF CANDIDATE GENE AND CONSTRUCTION OF LD MAPS

3.1 INTRODUCTION

Although the number of genes which may affect inter-individual warfarin dose is not known, this chapter will focus on 35 genes which were selected as the best candidates based on pharmacokinetic and pharmacodynamic information and were then tested for association to the two warfarin phenotypes: dose requirement and bleeding complication. Genes interacting with warfarin, function in vitamin K re-cycling or downstream pathways would be the most involved in dose variation or in increased bleeding risk.

Figure 3.1 summarises the genes selected in this study. A first set of 27 genes which does not include members of the protein disulfide isomerase (*PDI*) gene family, vitamin K epoxide reductase (*VKORC1*) and growth arrest-specific gene 6 (*GAS6*), were selected by Drs Mia Wadelius (Uppsala) and Munir Pirmohamed (Liverpool) [see (Wadelius and Pirmohamed 2007) for review]. The *VKORC1*, *GAS6* and *PDI* genes were included later in the project. These genes can be subdivided in five groups based on function:

(i) genes involved in transporting warfarin in the blood (*ORM1* and *ORM2*) and pumping warfarin out of liver cells (*ABCB1*) - may affect drug concentration in the blood and liver, and thereby resulting in dose variations.

(ii) enzymes metabolising warfarin (*CYP1A1*, *CYP1A2*, *CYP2C8*, *CYP2C9*, *CYP2C18*, *CYP2C19*, *CYP3A4* and *CYP3A5*). In addition to the genes encoding these cytochrome P450s, genes that regulate their expression (*NR1I2* and *NR1I3*) may also affect the required dose.

(iii) genes involved in intake and recycling of vitamin K (*APOE*, *VKORC1*, *GGCX*, *CALU*,

EPHX1, *NQO1*, *P4HB*, *PDIA2*, *PDIA3*, *PDIA4*, *PDIA5* and *PDIA6*) constitute another important group to target. The functional variants in these genes may dramatically influence the rate of reducing vitamin K and thereby a lower warfarin dose will be sufficient for effective anticoagulation.

(iv) a number of coagulation factors are activated by gamma-carboxylation. Polymorphisms in the genes encoding the vitamin K dependent (VKD) proteins *F2*, *F7*, *F9*, *F10*, *GAS6*, *PROC*, *PROS1*, and *PROZ* may well influence dose variation and blood coagulation may be easily toggle by low-dose warfarin due to impaired VKD proteins.

(v) some of the above genes are also regulated by *SERPINC1*. Finally, *F5* has been reported to be involved in both thrombosis and haemorrhage and therefore a strong candidate for bleeding complications.

The rational and detailed literature review for each gene is described below along with the results of the first step of this study which was to generate for each gene a comprehensive linkage disequilibrium map in a Swedish sample of 201 warfarin-treated patients enrolled at the Uppsala University by Dr Mia Wadelius and colleagues (see section 2.1.1 and Chapter IV).

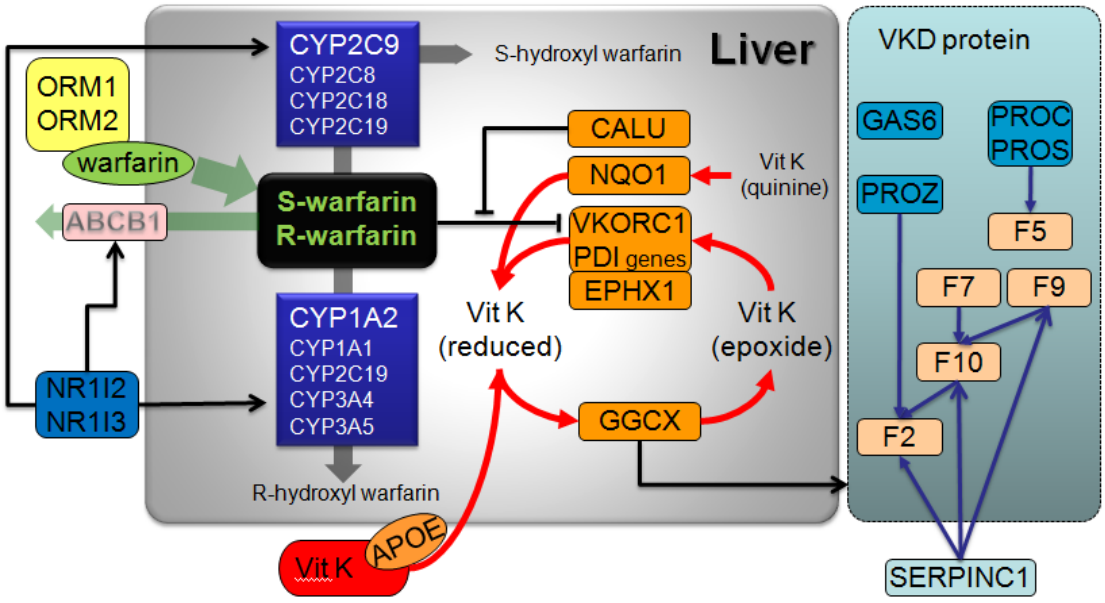


Figure 3.1. Genes selected in this study. (i) genes involved in transporting warfarin in the blood (*ORM1* and *ORM2*) and pumping warfarin out of liver cells (*ABCB1*); (ii) enzymes metabolising warfarin (*CYP1A1*, *CYP1A2*, *CYP2C8*, *CYP2C9*, *CYP2C18*, *CYP2C19*, *CYP3A4* and *CYP3A5*) and upstream regulated (*NR1I2* and *NR1I3*) ; (iii) genes involved in intake and recycling vitamin K (*APOE*, *VKORC1*, *GGCX*, *CALU*, *EPHX1*, *NQO1* and *PDI* genes); (iv) vitamin K dependent (VKD) proteins *F2*, *F7*, *F9*, *F10*, *GAS6*, *PROC*, *PROS1* and *PROZ*, and (v) other coagulation factors (*SERPINC1* and *F5*).

3.2 GENOTYPING STUDY DESIGN

The Uppsala cohort comprises 201 mainly Swedish Caucasian subjects (see section 2.1.1). At the start of this project, there was neither Hapmap data released, nor single nucleotide polymorphisms (SNPs) in public databases such as dbSNP with allele frequency information attached to them (or little, if there was any). Therefore, the first step was to establish a linkage disequilibrium (LD) map for each of the 35 selected genes in this study. We adopted SNPs as the genetic marker of choice to build LD maps because they are both abundant and their bi-allelic nature allows for high throughput genotyping techniques to be applied.

Initially, SNPs were randomly selected from the Ensembl database (http://www.ensembl.org) aiming at 5 kb spacing and minor allele frequency (MAF) equal or greater than 5%. The set of random SNPs was complemented with 73 non-synonymous SNPs and other candidate functional variants giving a total of 728 SNPs which passed assay design for the Sequenom platform (Table 3.1). In the first round of genotyping the 201 Swedish patients, we only obtained 189 common SNPs (MAF $\geq$ 5%) that passed study criteria. At the quality control step we removed SNP assays that failed clustering, those that were out of Hardy-Weinberg equilibrium (HWE, $p < 0.001$) or had a genotype call rate lower than 70% (Whittaker et al. 2005). We also disregarded SNPs with MAF less than 5%. To construct comprehensive LD maps we iteratively selected additional SNPs in gap regions defined physically, over 5 kb, or based on r^2 below 0.8. SNPs within gaps were extracted with a PERL script and followed by manual selection. SNP selection for filling in the outstanding gaps was based on the following criteria:

1. SNP with recorded allele frequency for Caucasians in dbSNP – two hit SNPs
2. SNP which were successfully genotyped in phase I Hapmap project.

3. SNP which had been genotyped internally at the Sanger Institute.

Table 3.1 shows the genotyping summary for each round of assay design. The use of HapMap and dbSNP in subsequent rounds of gap filling has greatly improved success rate from 26% to 43% as can be seen in Table 3.1. Note that in each round gaps are mostly difficult regions, such as highly homologous genes or within regions having repetitive elements, the validated SNP with allele frequency information bring greater opportunity from an initial 26% successful rate to ~43% for the 2nd and 3rd round.

Table 3.1. Summary of iterative genotyping.

Iteration	Attempted SNP	Informative (MAF ≥ 5%)	Disregard	%
1st	728	189	539	25.96%
2nd	202	86	116	42.57%
3rd	169	73	96	43.20%
4th	47	31	16	65.96%
subtotal	1146	379	767	33.07%

In the first three rounds all SNPs were assayed using Sequenom MassEXTEND technology whilst in the fourth round SNPs were assayed using iPLEX which allows higher multiplexing of markers. In the fourth round 6 genes of the protein disulfide isomerase A (PDI) family were examined. With available LD map information from Hapmap, informative SNPs for the six PDI genes were extracted to design the assays and therefore a much higher success assay ratio of 66% was achieved.

A total of 1146 SNPs were tested in order to construct LD maps with an r^2 above 0.8 in the Uppsala sample and over 90% genome coverage for each of the 35 candidate genes. Among

the 73 functional polymorphisms tested (http://www.cypalleles.ki.se/), most of them were excluded in subsequent analysis due to being monomorphic or having MAF lower than 5%. The criterion of 5% MAF was applied as a general threshold for common SNPs as in the HapMap project. Since we looked for common variants which influence warfarin dose, only informative SNPs were further analysed.

We identified 379 SNPs that passed assay quality control and the experimental criteria giving a good coverage for each candidate gene. In 767 disregarded SNP assays, 226 SNPs assays failed to pass the clustering quality control giving a failure rate of total designed SNP assays of 19.7 %. This is mainly because in multiplex experiments the presence of multiple oligo primers often leads to unpredictable behaviour including self priming. The latter causes a marker to show 'water calls' i.e. spurious genotypes in the absence of DNA template, which lower the confidence attached to all genotypes obtained with that assay In the 4^{th} assay design with 47 SNPs, a higher failure ratio of 25.5 % was seen as a result of the higher multiplex design used.

Table 3.2 shows the detailed results for each candidate gene. The number of SNPs which passed the initial clustering quality check (QC) step is counted in the QC column. Each QC SNP was then examined for Hardy-Weinberg equilibrium (HWE), genotype ratio (call rate) and MAF. The SNPs which failed the HWE test were due to high sequence similarity with other loci resulting in a non-specific amplification. Since the Hapmap information was used for the six protein disulfide isomerase genes, no rare or monomorphic SNPs were detected. No allele frequency information for a Swedish population was available at the start of the project but from previously described SNPs, 112 were found to be rare with MAF<5% and 293 could not be detected at all in this population.

Table 3.2. Summary of SNP genotyping for each of the candidate genes.

			MAF ≤ 5%		Failed	
Gene	QC	polyM	rare	monoM	HWE	call rate (70%)
ABCB1	113	38	34	25	3	13
APOE	3	2		1		
CALU	14	9		5		
CYP1A1	15	3	4	7		1
CYP1A2	23	3	7	13		
CYP2C18	22	14	4	3	1	
CYP2C19	35	10	3	16	2	4
CYP2C8	18	12	2			4
CYP2C9	36	19	2	12	1	2
CYP3A4	45	2	6	30	2	5
CYP3A5	24	7	6	8		3
EPHX1	42	25	5	6	2	4
F10	45	15	6	18		6
F2	22	10	3	8		1
F5	74	41	5	19	1	8
F7	32	11	2	15		4
F9	29	11	1	16		1
GAS6	5	4				1
GGCX	16	9	1	4		2
NQO1	18	9	1	7		1
NR1I2	41	20	6	8	2	5
NR1I3	19	9	3	5	1	1
ORM1, & ORM2	29	6		10	8	5
P4HB	6	5				1
PDIA2	4	4				
PDIA3	4	4				
PDIA4	8	7			1	
PDIA5	6	5			1	
PDIA6	7	6			1	
PROC	25	13		5		7
PROS1	50	11	3	20	7	9
PROZ	27	13	3	10		1
SERPINC1	23	9	1	8	1	4
VKORC1LD	39	13	4	13		9
	919	379	112	292	34	102

3.3 CANDIDATE GENES IN WARFARIN TRANSPORTATION

Three genes involved in transporting warfarin in the blood (*ORM1*, *ORM2*) and pumping warfarin out of liver cells (*ABCB1*) were selected (Table 3.3). *ORM1* and *ORM2* are located back to back on chromosome 9q. Both genes have 6 exons, encode peptides comprising 201 amino-acid residues and have very high DNA sequence homology (Figure 3.2). The latter caused extreme difficulty in designing genotyping assays for exonic SNPs. *ABCB1* spans 209 kb on chromosome 7 and belongs to the ATP-binding cassette transporter gene family encoding various cellular pumps in different cells. *ABCB1* is particularly a cellular efflux pump for xenobiotics.

Table 3.3. Candidate genes in warfarin transportation.

Gene	Location	Exons	Transcript	Protein (residues)	Function of protein
ORM1	Chr 9: 114083890 - 114087309 bp	6	802 bps	201	A plasma glycoprotein that functions as a carrier of warfarin in the blood
ORM2	Chr 9: 114171703 - 114175086 bp	6	760 bps	201	A plasma glycoprotein that functions as a carrier of warfarin in the blood
ABCB1	Chr 7: 85668428 - 85877818 bp	29	4643 bps	1279	A cellular efflux pump for xenobiotics.

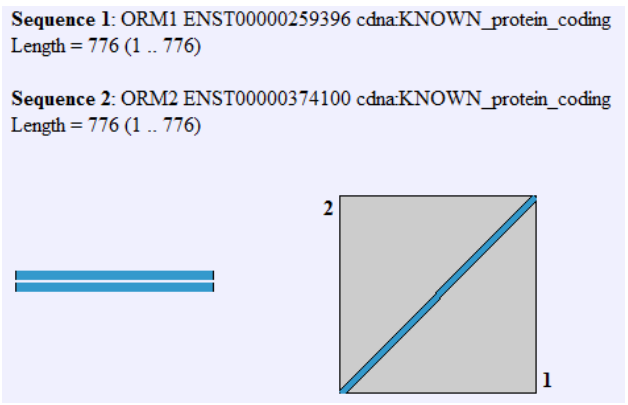


Figure 3.2. cDNA sequence alignment of ORM1 and ORM2. cDNA sequences of both genes contains 776 nucleotides, and only 31 nucleotides are different resulting in a 96% similarity.

3.3.1 ORM1 and ORM2

Warfarin has been intensively studied over decades as an effective anticoagulant. When warfarin is taken, it will be quickly and nearly completely absorbed from the stomach and the upper gastrointestinal tract and then bound to serum albumin and alphal-acid glycoprotein encoded by *ORM1* (orosomucoid 1) and *ORM2* (orosomucoid 2). In plasma, ORM proteins are presented as a mixture of ORM1 and ORM2 at a molar ratio of 3:1 (Yuasa et al. 1997).

A chromatographic study (Nakagawa et al. 2003) has shown that warfarin has different affinity to orosomucoid protein; harbouring amino acid changes due to genetic variation which suggested that orosomucoid protein might influence warfarin dose. Meanwhile, serum albumin has very robust transporting functionality for a broad range of proteins and thus, it is less likely to play a role in inter-individual dose variability. No evidence was reported in the literature for an association between albumin and warfarin dose.

As described earlier, due to high sequence homology (96%), 28 SNPs in *ORM1, ORM2* and flanking regions were tested, but only 6 SNPs located in either intronic or flanking regions gave good results (Table 3.4). None of these 6 SNPs have been reported to be functionally important. LD analysis indicated that the 6 SNPs are in strong LD (Figure 3.3A). Hapmap project phase II data (CEU panel) also show that both *ORM1* and *ORM2* are located at the end of an LD block (Figure 3.3B). In this chapter the phase II result of Hapmap project was compared with Swedish LD structure for each of the candidate genes.

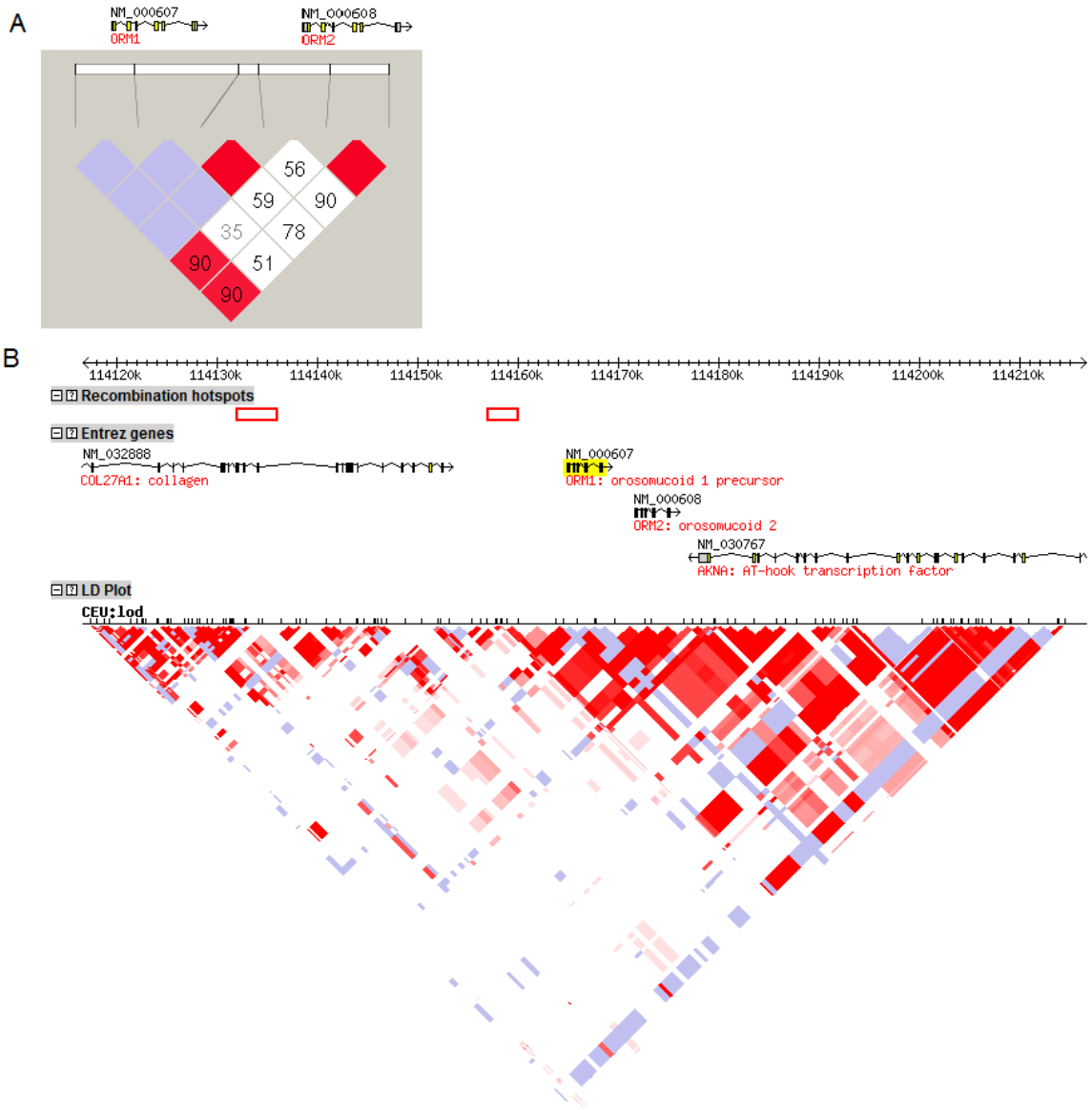


Figure 3.3. (A) Swedish LD structure; (B) LD structure from Hapmap project phase II result. (A) None of exonic SNPs in *ORM1* and *ORM2* were successfully genotyped because of high cDNA sequence similarity (96%). (B) LD structure from Hapmap phase II result indicates the *ORM1* and *ORM2* are at the end of a LD block, and a recombination hot spot is 5 kb away.

Table 3.4. Genotyping summary of *ORM1* and *ORM2*.

Genes	Study SNPs (tested, passed)
ORM1	15 tested, 3 passed study criteria
Coding ns	0
Coding s	0
UTR	0
Intronic	1: rs10982151
Flanking	2: rs2787337, rs1687390
ORM2	13 tested, 3 passed study criteria
Coding ns	0
Coding s	0
UTR	0
Intronic	2: rs17230081, rs1976193
Flanking	1: rs3762055

3.3.2 ABCB1 (MDR1)

Preliminary evidence has demonstrated that P-glycoprotein transports warfarin (as a substrate) through liver plasma membranes using an inhibition assay (Sussman N 2002). This P-glycoprotein is encoded by *ABCB1* (ATP-binding cassette transporter B1) and is also named *MDR1* (multidrug resistance protein 1). Increased expression of *ABCB1* in the small intestine will reduce the absorption of drugs which are substrates for P-glycoprotein, and thus result in a reduced bioavailability and unattained therapeutic plasma concentration of a drug (Cascorbi et al. 2001).

There are a number of reports showing that sequence variants in *ABCB1* influence its transcriptional and translational expression, as well as the pharmacokinetic profiling of various drugs (Ishikawa et al. 2004). A study demonstrated that a haplotype bearing the C3435T variant in exon 26 was over-represented among patients requiring a lower maintenance dose of warfarin (Wadelius et al. 2004). Coincidently, C3435T variation has been shown to reduce drug efflux which is in agreement with the need of a lower dose because of reduced removal of toxic metabolites from the liver cells. There is also another possibility according to which the increased intestinal expression of P-glycoprotein can reduce the absorption of warfarin resulting in a higher maintenance dose (Cascorbi et al. 2001). However, a recent study demonstrated that C3435T affects the timing of co-translational folding and insertion of P-glycoprotein into the membrane (Kimchi-Sarfaty et al. 2007). The observed lower level of membrane insertion of the T variant would be consistent with lower efflux of warfarin from liver cells.

In this study, 113 SNPs in *ABCB1* passed QC and 38 with MAF higher than 5% were

informative. None of these 38 SNPs had been previously reported to be functionally important, however, two nsSNPs cause a change of amino acid composition (N21D and A893S/T) and one nsSNP may even change the translational initiation (Table 3.5).

The genomic structure of *ABCB1* and corresponding LD map is shown in Figure 3.4. Panel A in Figure 3.4 shows the LD map in the Swedish population sample from Uppsala in which the gene seems to be split in the middle into two LD blocks. The same is found in Hapmap (Figure 3.4B) where LD decays as a function of rapidly evolving SNPs in the first intron. Two recombination hot spots are found in Hapmap: one in the centre of *ABCB1* genomic region and the other one at the 3' end of the gene (Figure 3.4).

Table 3.5. Genotyping summary of *ABCB1*.

Genes	Study SNPs (tested, passed)
ABCB1	113 tested, 38 passed study criteria
Coding ns	3: rs2214102 (changes translation initiation), rs9282564 (N21D), rs2032582 (A893S/T)
Coding s	1: rs1045642
UTR	1: rs3842
Intronic	33: rs2188531, rs6465117, rs17328991, rs10267099, rs2157926, rs2214101, rs17149824, rs4728709, rs9282564, rs1858923, rs3789243, rs1202181, rs1202172, rs1989830, rs1202179, rs1202180, rs10260862, rs2235015, rs1202167, rs1202169, rs955000, rs868755, rs1922240, rs2235033, rs2235035, rs2235013, rs2091766, rs2235046, rs1922242, rs4148737, rs2235040, rs6959435, rs4148742, rs2235067

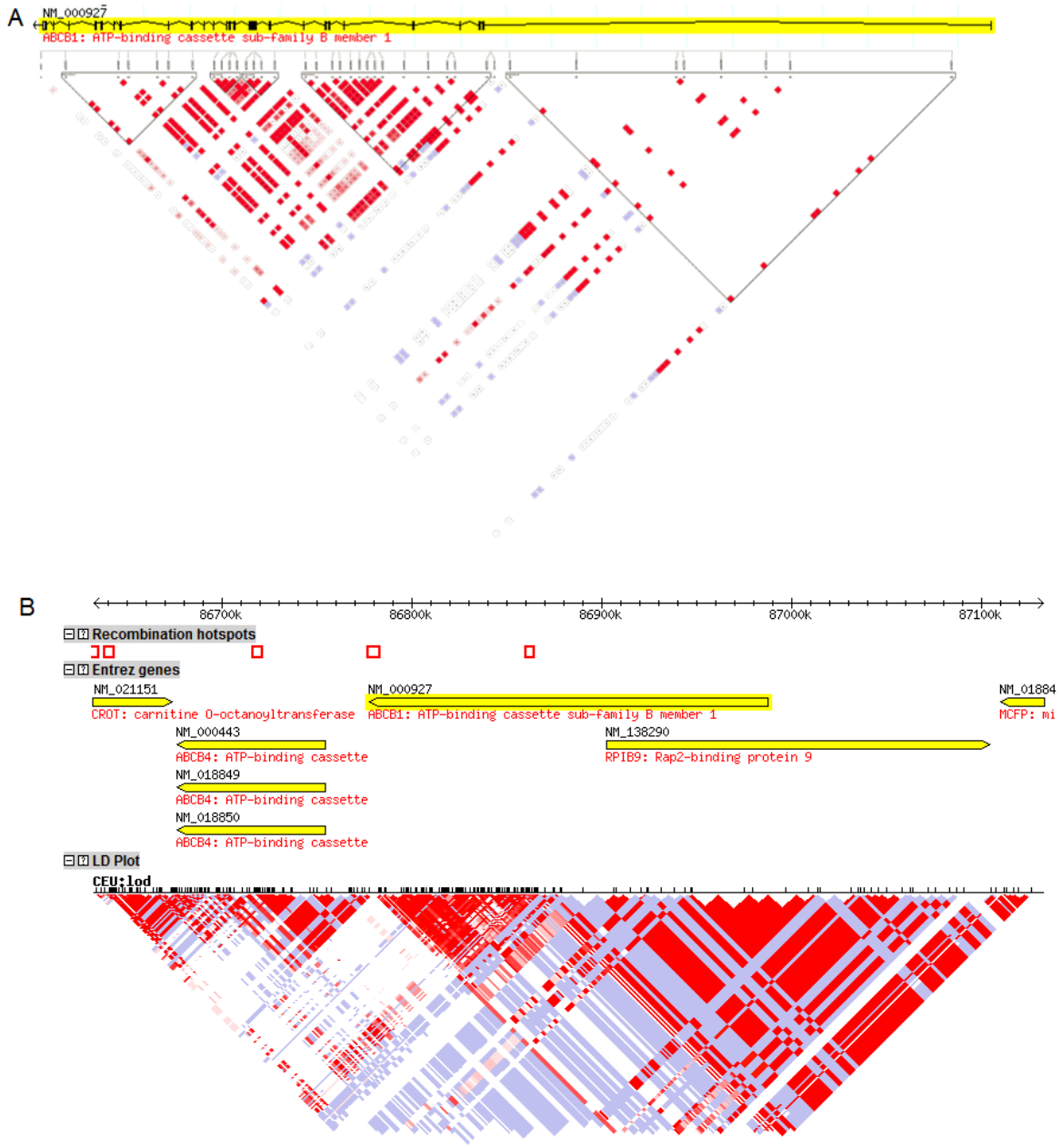


Figure 3.4. LD structure from (A) 201 Swedish and (B) Hapmap CEU. (A) LD in Swedish population suggests a recombination hot spot in the centre of ABCB1. (B) This hot spot is also discovered in HapMap Caucasian population.

3.4 CANDIDATE GENES IN WARFARIN METABOLISM

Warfarin consists of a racemic mixture of two active optical isomers - R and S forms. Each is cleared by different pathways (discussed later this section). S-warfarin has five times the potency of the R-enantiomer with respect to vitamin K antagonism.

Table 3.6 lists the cytochrome P450 metabolising enzymes which convert warfarin to less toxic forms. The *CYP2C9, CYP2C8, CYP2C18, and CYP2C19* genes are clustered in a region of high linkage disequilibrium on chromosome 10 (Ahmadi et al. 2005), whereas *CYP1A1* and *CYP1A2* are located back to back on chromosome 15 and *CYP3A4* and *CYP3A5* on chromosome 7.

Table 3.6. Candidate genes in warfarin metabolism.

Gene	Location	Exons	Transcript	Protein (residues)	Function of protein
CYP2C9	Chr 10: 96688405 - 96739137 bp	9	1847 bps	490	Polymorphic hepatic drug metabolising enzyme. Metabolism of S-warfarin
CYP2C8	Chr 10: 96786520 - 96819244 bp	9	1923 bps	490	Polymorphic hepatic drug metabolising enzyme. Minor pathway for R & S-warfarin
CYP2C18	Chr 10: 96432700 - 96485937 bp	9	2418 bps	490	Found in the liver and lung. Minor pathway for R & S-warfarin
CYP2C19	Chr 10: 96437901 - 96603007 bp	9	1901 bps	490	Polymorphic hepatic drug metabolising enzyme. Minor pathway for R & S-warfarin
CYP1A1	Chr 15: 72798943 - 72804930 bp	7	2601 bps	512	Extrahepatic oxidation, inducible. Metabolism of R-warfarin
CYP1A2	Chr 15: 72828257 - 72834505 bp	7	1618 bps	516	Hepatic oxidation, inducible. Metabolism of R-warfarin
CYP3A4	Chr 7: 97889181 - 97916385 bp	13	2768 bps	503	Hepatic oxidation, inducible. Metabolism of R-warfarin
CYP3A5	Chr 7: 97780394 - 97812183 bp	13	1707 bps	502	Polymorphic hepatic and extrahepatic oxidation. Metabolism of R-warfarin

3.4.1 S-warfarin metabolism (CYP2C8, CYP2C9, CYP2C18, and CYP219)

When warfarin circulates to the liver, the S-enantiomer is mainly converted to 6- and 7-hydroxywarfarin by CYP2C9 whilst CYP2C8 and CYP2C19 metabolise small amounts of S-warfarin to the 4-hydroxyl metabolite (Kaminsky and Zhang 1997).

There are a number of non-synonymous coding variants identified in *CYP2C9* with differential, mostly decreased, enzymatic activities (Schwarz 2003). Some of these variants were reported in particular ethnic populations including the *4 allele which is identified in Japanese, *5 and *6 alleles identified in African-American, and *11 which is relatively rare in European Caucasian and African-American (Schwarz 2003).

Among the 30 reported variants which are functionally important, the *2 and *3 variants have been convincingly shown to be associated with reduced warfarin dose (Aithal et al. 1999; Higashi et al. 2002; Wadelius et al. 2005; Wadelius et al. 2004). Compared to a CYP2C9 metaboliser who is homozygous for the *1 wild type allele, a homozygote for the *2 allele will have a reduced CYP2C9 enzymatic activity of 12%, whereas a homozygote for the *3 allele has only 5% metabolising activity. Many studies have shown that patients with *2 and *3 alleles require lower daily warfarin dose (Aithal et al. 1999; Higashi et al. 2002; Wadelius et al. 2005; Wadelius et al. 2004). A meta-analysis of nine previous studies indicated that the *2 allele would reduce dose requirement by 17% whereas *3 allele comprehensively lowered the required dose by 37% (Sanderson et al. 2005).

Of the genes listed in Table 3.6 only *CYP2C9* has been reported to be associated with warfarin dose. Although many functional variants have been reported in these CYP genes,

only a few of them were found to be polymorphic in the 201 Swedish patients. Table 3.7 lists all informative SNPs including functional variants used for analysis. The LD structure of these four genes is discussed in the chapter 4 (section 4.4) together with the association analysis for warfarin dose requirement. Apart from those alleles described on the CYP website (http://www.cypalleles.ki.se/), SNP rs2281891 (T385M) in *CYP2C18* has been reported to be associated with differential turnover rate for drugs (Goldstein et al. 1994; Kaminsky et al. 1993) and is polymorphic in this Swedish cohort.

Table 3.7. Genotyping summary for candidate genes in S-warfarin metabolism.

Genes	Study SNPs (tested, passed)	SNP aliases
CYP2C9	36 tested, 19 passed study criteria	rs1799853 (formerly rs17110268) = CYP2C9*2, coding R144C, low activity; rs1057910 = CYP2C9*3, coding I359L, very low activity
Coding ns	2 : rs1057910 (I359L), *2 rs1799853 (R144C)	
Coding s	1: rs1057911	
UTR	0	
Intronic	14: rs2298037, rs9332222, rs9332214, rs9332197, rs1934966, rs1934964, rs9325473, rs1856908, rs4917639, rs2153628, rs2475376, rs10509679, rs2860905, rs9332108	
Flanking	2: rs4917636, rs4607998	
CYP2C8	18 tested, 12 passed study criteria	rs17110453 = CYP2C8*1C, 5' upstream, function unknown; rs11572080 = CYP2C8 *3, coding R139K, decreased activity; rs1058930 = CYP2C8 *4, coding I264M, decreased activity
Coding ns	2: rs11572080 (R139K), rs1058930 (I264M)	
Coding s	0	
UTR	1: rs1058932	
Intronic	7: rs2275622, rs3752988, rs1341163, rs947173, rs1891071, rs2275620, rs7898759	
Flanking	2: rs1557044, rs17110453	
CYP2C18	22 tested, 14 passed study criteria	rs2281891= coding T385M, high metabolism of certain substrates, but not warfarin
Coding ns	1: rs2281891 (T385M)	
Coding s	0	
UTR	1: rs2860840	
Intronic	10: rs10509675, rs7919273, rs1926711, rs7898763, rs7099637, rs7896133, rs7478002, rs2901783, rs2860837, rs1926706	
Flanking	2: rs10736086, rs12249418	
CYP2C19	35 tested, 10 passed study criteria	rs17879456 = rs4244285, CYP2C19*2A, splicing defect, no enzyme activity; rs17882687 = CYP2C19*15, coding I19L, effect on enzyme unknown
Coding ns	2: rs17882687 (I19L), rs17879456 (splicing defect)	
Coding s	1: rs3758580	
UTR	0	
Intronic	4: rs1853205, rs4244284, rs4417205, rs17882419	
Flanking	3: rs12248560, rs3814637, rs4250786	

3.4.2 R-warfarin metabolism (CYP1A1, CYP1A2, CYP3A4 and CYP3A5)

The R-enantiomer is mainly metabolised to 6-, 8-, and 10-hydroxylated form. The minor metabolites include 4- and 7-hydroxywarfarin. The 6- and 8-hydroxyl metabolites are formed predominately by CYP1A2 with some contributions from CYP1A1 and CYP2C19 (Kaminsky and Zhang 1997). The 10-hydroxyl form is generated by CYP3A of which activity is derived from CYP3A4 and CYP3A5 with similar substrate specificities (Kaminsky and Zhang 1997; Kuehl et al. 2001). The genotyping results are summarised in Table 3.8.

Table 3.8. Genotyping summary for candidate genes in R-warfarin metabolism.

Genes	Study SNPs (tested, passed)	SNP aliases
CYP1A1	15 tested, 3 passed study criteria	-
Coding ns	0	
Coding s	0	
UTR	0	
Intronic	2: rs4646421, rs2606345	
Flanking	1: rs2470893	
CYP1A2	23 tested, 3 passed study criteria	rs2470890: CYP1A2*1B, coding synonymous, function unknown; rs762551 = CYP1A2*1F, 5'UTR, higher inducibility
Coding ns	0	
Coding s	1: rs2470890	
UTR	1: rs762551	
Intronic	1: rs2472304	
Flanking	0	
CYP3A4	45 tested, 2 passed study criteria	rs11773597 = CYP3A4*1F, 5'upstream, function unknown
Coding ns	0	
Coding s	0	
UTR	0	
Intronic	1: rs11773597	
Flanking	1: rs2242480	
CYP3A5	24 tested, 7 passed study criteria	rs776746 = CYP3A5*3, splicing defect, low activity; g-3844G>A, 5'upstream, function unknown; rs28365067 = intronic g5215C>T, function unknown; rs28365094 = intronic g27050A>G, function unknown
Coding ns	1: rs776746 (splicing defect)	
Coding s	0	
UTR	0	
Intronic	3: rs6976017, rs28365067, rs28365094	
Flanking	3: rs4646457, g-3844G>A (not in dbSNP), rs15524	

The LD map of *CYP1A1* and *CYP1A2* is shown in Figure 3.5A. The poor genotyping coverage in this region is not resulting from high sequence similarity or repetitive regions

(Figure 3.6). In fact, typing of 38 SNPs was attempted across this region and 31 SNPs were discovered which are either rare (11 SNPs, MAF ≤ 5%) or monomorphic (20 SNPs). This is in complete agreement with Hapmap phase II data which report only 7 SNPs in the two genes and 5 kb upstream and downstream region (Figure 3.5B). These results suggest that the CYP1A locus does not tolerate the accumulation of sequence variants which may reflect functional constraints.

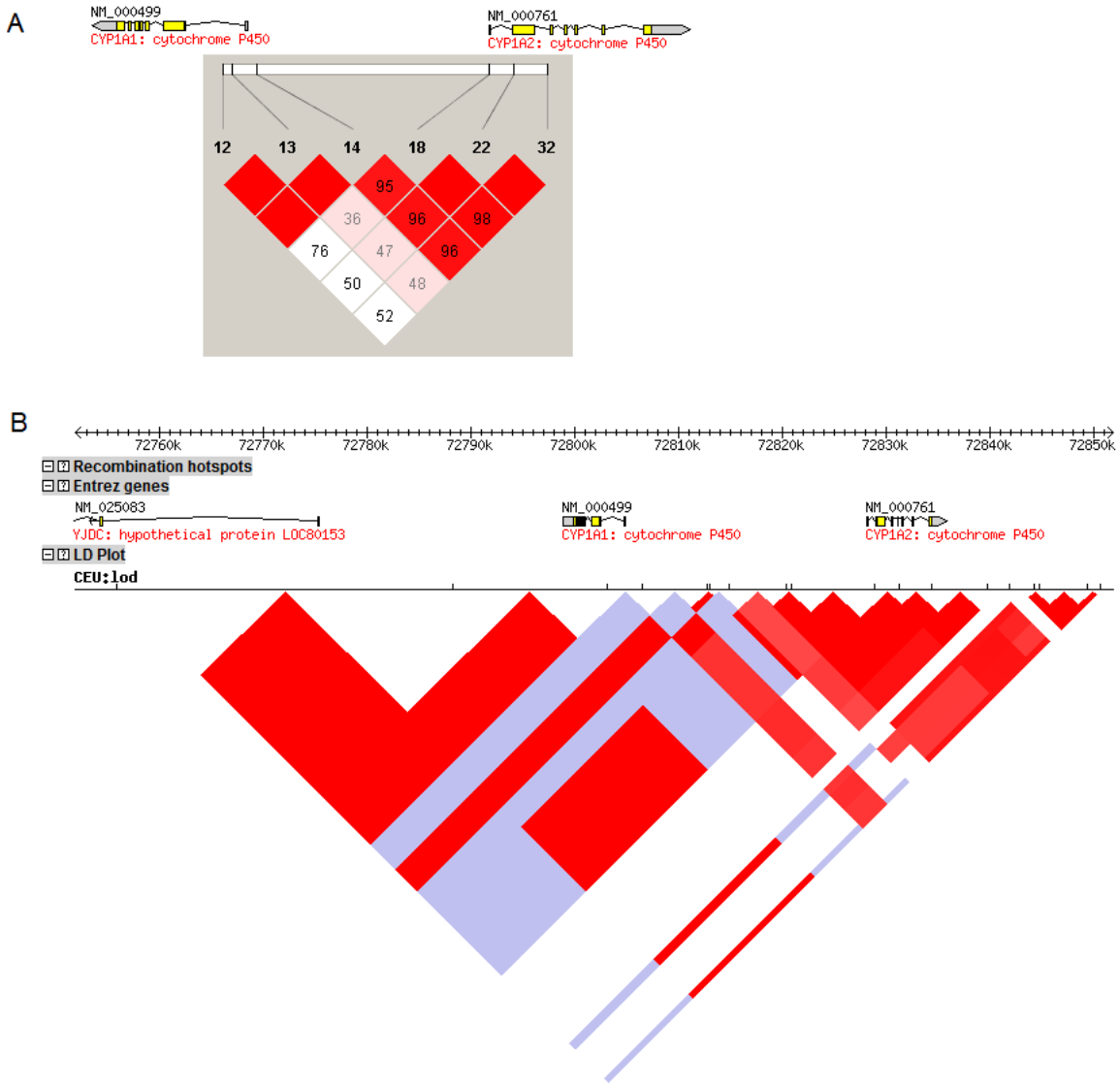


Figure 3.5. LD structure of CYP1A1 and CYP1A2 in (A) Swedish and (B) Hapmap CEU populations.

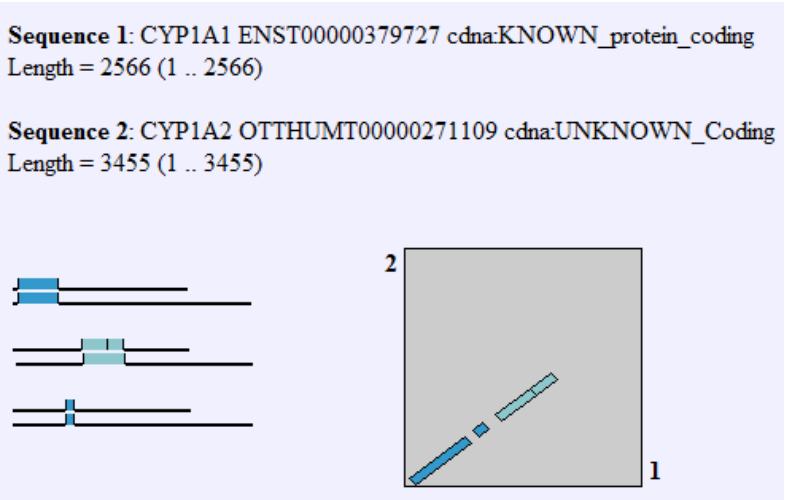


Figure 3.6. Pairwise cDNA sequence alignment of *CYP1A1* and *CYP1A2*. Three segments are high homologous with 88%, 74%, 81%, respectively in the first two third of the genes.

Figure 3.7 shows the LD structures of *CYP3A4* and *CYP3A5* in the Swedish and Hapmap Caucasian sample. After attempting two rounds, only two SNPs were obtained from the last intron and 3' flanking region of *CYP3A4*. A total of 45 SNPs were tested in *CYP3A4* with 6 SNPs found to be rare and 30 monomorphic (Table 3.2). A sparse coverage is also seen in the Hapmap data which show the presence of a recombination hot spot in the centre of *CYP3A4* gene (Figure 3.7B).

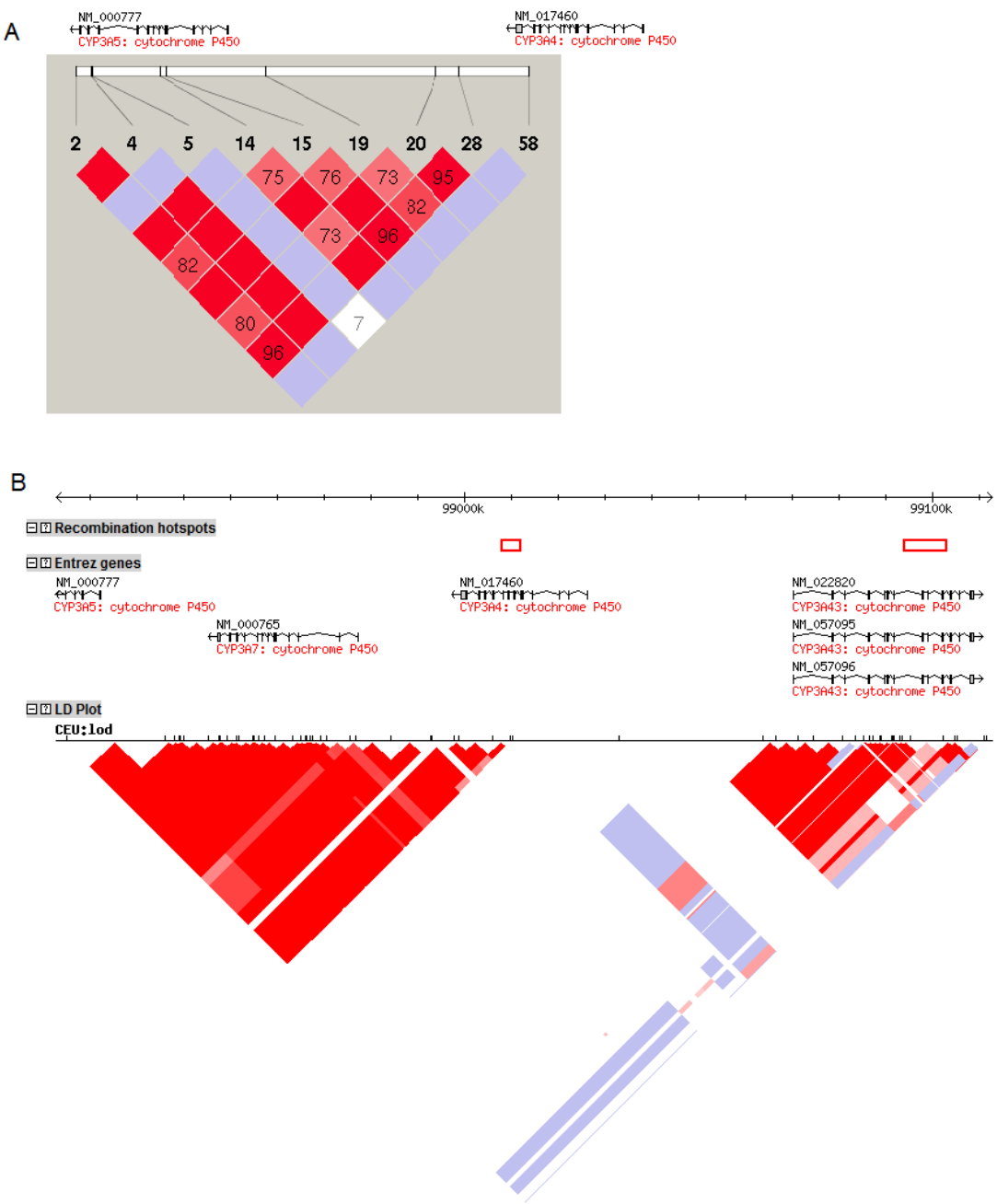


Figure 3.7. LD structure of *CYP3A4* and *CYP3A5* in (A) Swedish and (B) Hapmap CEU populations.

3.4.3 P450 inducibility

Since much of the detoxification of xenobiotic compounds is performed by CYP enzymes, it is also important to look at the genes that regulate CYP expression and activity. Some studies

have shown that these CYP enzymes are inducible especially warfarin is prescribed with other drugs (Lehmann et al. 1998; Moore et al. 2000). In that respect, two genes *NR1I2* and *NR1I3* were identified as regulating the expression of CYP genes (Table 3.9) and therefore selected for analysis as they may harbour variants that influence warfarin dose.

Table 3.9. Candidate genes of regulating warfarin metaboliser.

Gene	Location	Exons	Transcript	Protein (residues)	Function of protein
NR1I2	Chr 3: 120982021 - 121020021 bp	9	2753 bps	473	Mediates induction of CYP2C9, CYP3A4, other CYP enzymes and ABCB1
NR1I3	Chr 1: 158012528 - 158021028 bp	9	1337 bps	348	Transcriptional regulation of a number of genes including CYP2C9 andCYP3A4

The pregnane X receptor (PXR) encoded by *NR1I2* (nuclear receptor subfamily 1, group I, member 2) is activated by a variety of endogenous and exogenous chemicals including St John's Wort; a herbal antidepressant which interacts with warfarin (Table 3.10). PXR is also reported to induce CYP2C9, CYP3A4, and other CYPs (Chen et al. 2004; Lehmann et al. 1998). Interestingly, PXR also regulates the drug efflux by activating expression of ABCB1 (MDR1) (Geick et al. 2001; Synold et al. 2001).

The constitutive androstane receptor (CAR) which is encoded by *NR1I3* (nuclear receptor subfamily 1, group I, member 3), is a nuclear hormone receptor that functions cooperatively with PXR (*NR1I2*) to detoxify xenobiotics (Table 3.10). CAR is also reported to be associated with induced transcription of *CYP2C9* and *CYP3A4* (Assenat et al. 2004).

Table 3.10. Genotyping summary of *NR1I2* and *NR1I3*.

Genes	Study SNPs (tested, passed)
NR1I2	41 tested, 20 passed study criteria
Coding ns	0
Coding s	0
UTR	5: rs3814057, rs1054191, rs3732360, rs3732359, rs1523127
Intronic	13: rs2472682, rs3732357, rs3732356, rs1464602, rs7643645, rs2461818, rs13059232, rs2461823, rs2472677, rs1403527, rs2056530, rs2472672, rs2276706
Flanking	2: rs1523130, rs7643038
NR1I3	19 tested, 9 passed study criteria
Coding ns	0
Coding s	1: rs2307424
UTR	0
Intronic	5: rs2502804, rs6686001, rs3003596, rs2307418, rs4073054
Flanking	3: rs2501870, rs7530560, rs4233368

20 SNPs in *NR1I2* and 9 SNPs in *NR1I3* were analysed (Figure 3.8). The first exon of *NR1I2* is un-translated and two recombination hot spots map in the first and second intron respectively (Figure 3.8A). Hapmap results confirm the presence of the two recombination hot spots in Caucasians which are located 20 kb apart (Figure 3.9). This suggests a higher mutation rate in this gene, and a lot of SNPs have been reported in Ensembl SNP database. In contrast to *NR1I2*, SNPs in *NR1I3* suggest a strong LD across this region which is in agreement with Hapmap (Figure 3.8B).

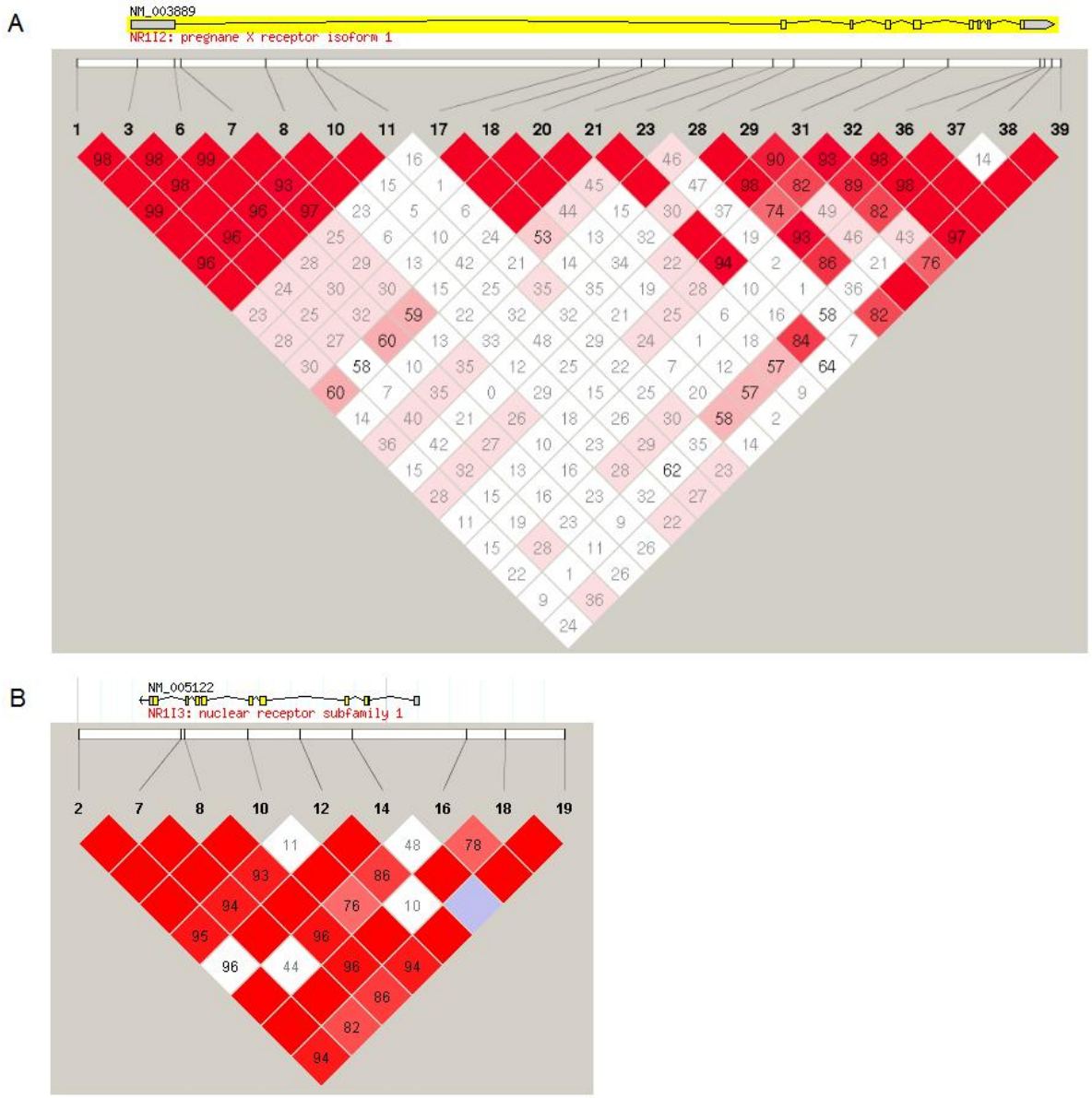


Figure 3.8. LD structure of (A) *NR1I2* and (B) *NR1I3*. (A) The transcription of *NR1I2* starts from the second exon, and the whole exon 1 is untranslated. Two potential recombination hot spots in intron 1 and intron 2 were observed in *NR1I2*. (B) SNPs in and flank on *NR1I3* indicate a strong LD in this region.

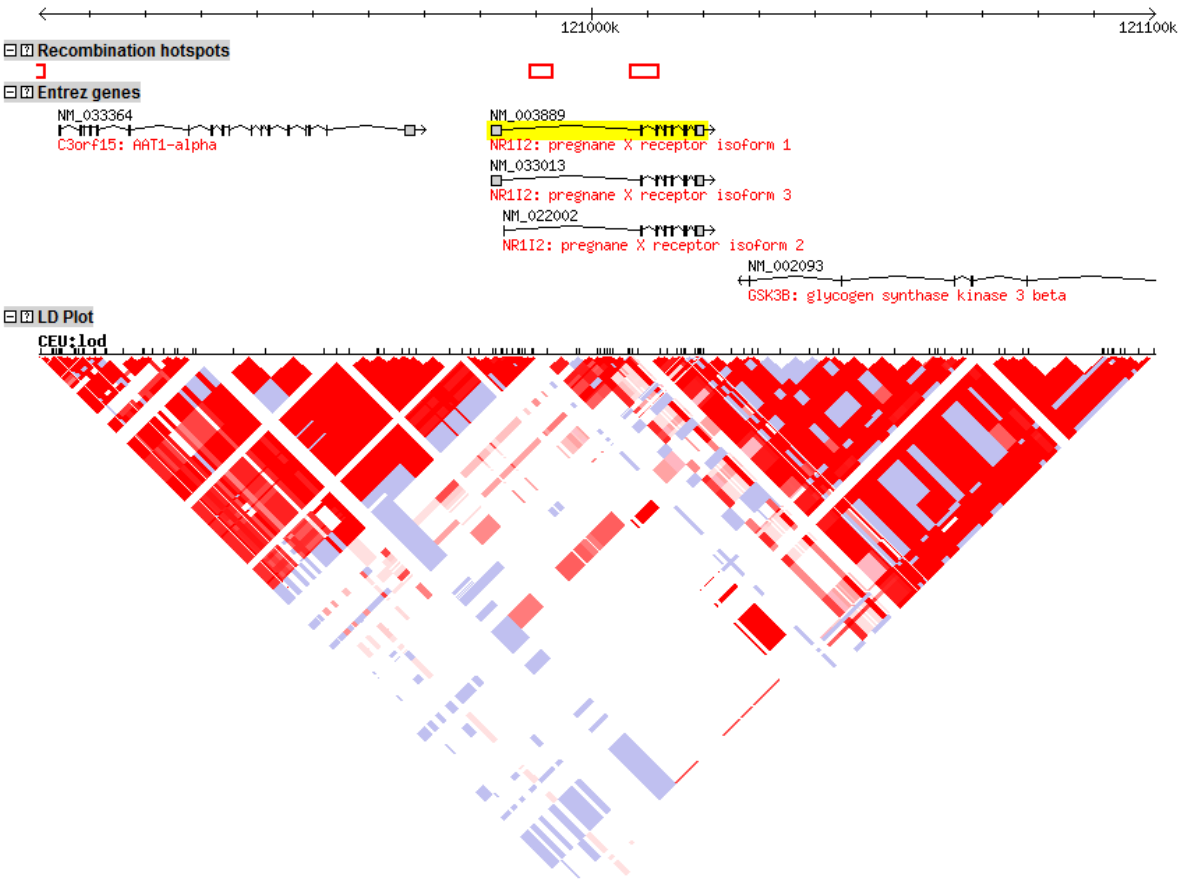


Figure 3.9. Hapmap CEU result for *NR1I2* region. Two recombination hot spots were identified in *NR1I2* in 96 Caucasians. This suggests a potential chromosome instability and higher mutation rate in this gene.

3.5 CANDIDATE GENES IN VITAMIN K INTAKE AND RECYCLING

Vitamin K is an important cofactor to activate certain blood coagulation factors. Warfarin acts on the vitamin K epoxide reductase (VKOR) complex on the endoplasmic reticulum (ER) membrane in liver cells to toggle the recycling of vitamin K and thereby, is interfering with blood coagulation. A diet with high vitamin K update would prevent the action of warfarin and result in a higher maintenance dose. Nowadays, more and more people try to live healthier eating more leafy green vegetables, Brassica vegetables, and fruits which are the major sources of daily vitamin K1 uptake. Table 3.11 lists all genes known to be involved in the recycling of vitamin K1

Table 3.11. Candidate genes in vitamin K intake and recycling.

Gene	Location	Exons	Transcript	Protein (residues)	Function of protein
APOE	Chr 19: 50100879 - 50104489 bp	4	1179 bps	317	Apolipoprotein E serves as a ligand for receptors that mediate the uptake of vitamin K
VKORC1	Chr 16: 31009677 - 31013777 bp	3	997 bps	163	A hepatic epoxide hydrolase that catalyses the reduction of vitamin K. The target of warfarin
EPHX1	Chr 1: 222304587 - 222339995 bp	9	1605 bps	455	A hepatic epoxide hydrolase with the potential to reduce vitamin K
NQO1	Chr 16: 68300807 - 68317893 bp	6	2448 bps	274	A detoxifying enzyme that has the potential to reduce the quinine form of vitamin K
GGCX	Chr 2: 85687865 - 85700237 bp	15	3155 bps	758	Carboxylates vitamin K dependent coagulation factors and proteins in the vitamin K cycle
CALU	Chr 7: 127973368 - 128005478 bp	7	3316 bps	315	Binds to the vitamin K epoxide reductase complex and inhibits the effect of warfarin
P4HB	Chr 17: 77,394,322-77,411,834 bp	11	2580 bps	508	Providing electrons to reduce CXXC centre in VKORC1
PDIA2	Chr16: 273,153-277,216 bp	11	1698 bps	525	Providing electrons to reduce CXXC centre in VKORC1
PDIA3	Chr15: 41,825,882-41,852,769 bp	13	3727 bps	505	Providing electrons to reduce CXXC centre in VKORC1
PDIA4	Chr7: 148,331,087-148,356,666 bp	10	2903 bps	645	Providing electrons to reduce CXXC centre in VKORC1
PDIA5	Chr3: 124,268,599-124,363,641 bp	16	1656 bps	262	Providing electrons to reduce CXXC centre in VKORC1
PDIA6	Chr2: 10,840,968-10,870,421 bp	13	2338 bps	440	Providing electrons to reduce CXXC centre in VKORC1

3.5.1 APOE

Vitamin K1 is a fat-soluble vitamin and is absorbed from the small intestine, together with dietary fat. It is then circulated in the blood by chylomicrons, large lipoprotein molecules secreted by absorptive cells of the small intestine. Vitamin K1 conjugated chylomicron becomes matured by acquiring apolipoprotein C-II (APOC2) and apolipoprotein E (APOE) from High Density Lipoproteins (HDL) and is subsequently cleared by the liver through an APOE receptor specific uptake (Berkner and Runge 2004; Lamon-Fava et al. 1998). It has been reported that different APOE alleles lead to variable uptake efficiency of the chylomicron with evidence showing that the *E4 allele is faster at uptake than *E3 and which is then faster than *E2 (Kohlmeier et al. 1996). Patients carrying the APOE*E2 allele, which is allegedly associated with less efficient uptake of chylomicron and thereby vitamin K1, had an increased risk of warfarin associated intracerebral haemorrhage (Rosand et al. 2000). Table 3.12 shows the two SNPs, rs429358 and rs7412, which passed study criteria and are sufficient to discriminate the APOE haplotypes carrying the E2, E3, and E4 allele.

Table 3.12. Genotyping summary of *APOE*.

Genes	Study SNPs (tested, passed)	SNP aliases
APOE	3 tested, 2 passed study criteria	rs429358 = coding C130R
Coding ns	2: rs429358 (C130R), rs7412 (R176C),	(formerly C112R); rs7412 =
Coding s	0	coding R176C (formerly R158C) ;
UTR	0	These 2 SNPs discriminate
Intronic	0	between the haplotypes E2, E3,
Flanking	0	E4

3.5.2 VKORC1

Warfarin and other vitamin K antagonists preclude the regeneration of vitamin K from its

oxidative forms by targeting VKOR. The gene encoded for VKOR was poorly understood up to 2004, when the coding gene was identified by two studies which demonstrated that the vitamin K epoxide reductase subunit 1 (VKORC1) is the direct target of warfarin by RNAi inhibition and functional assay (Li et al. 2004; Rost et al. 2004a). *VKORC1* is a small gene located on chromosome 16 which comprises 3 exons and encodes a peptide of 163 amino acids. Although this gene was shown to be the target for warfarin, the mechanism by which it functions as a reductase remains unclear.

Twenty-five publicly available SNPs were tested and only four polymorphic SNPs were obtained in our Swedish sample; two intronic SNPs (rs9934438 and rs2359612), one in the 3'-UTR (rs7294), and one 5' upstream (rs9923231) (Table 3.13). According to Hapmap, *VKORC1* is located near one end of a large LD block in Caucasians (CEU panel; Figure 3.10). In later stages of the study a further set of 14 SNPs, lying outside *VKORC1*, were selected to analyse this large LD block. For illustration purposes the genomic structure of *VKORC1* and LD architecture of this locus is discussed alongside the results of the association analysis in chapter 4, section 4.3.1 and 4.3.2).

Table 3.13. Genotyping summary of *VKORC1* and its nearby flanking region.

Genes	Study SNPs (tested, passed)	SNP aliases
VKORC1	39 tested, 13 passed study criteria	rs9923231 = upstream 3673 or -1639 G>A, low expression?; rs9934438 = intronic 6484 or 1173 C>T, function unknown; rs2359612 = intronic 7566 or 2255 C>T, function unknown; rs7294 = 3'UTR 9041 or 3730 G>A, function unknown
Coding ns	0	
Coding s	0	
UTR	1: rs7294	
Intronic	2: rs9934438, rs2359612	
Flanking	10: rs4889537, rs9923231, rs8046978, rs4889599, rs11642603, rs4889630, rs7405035, rs4889490, rs11642466, rs7194347	

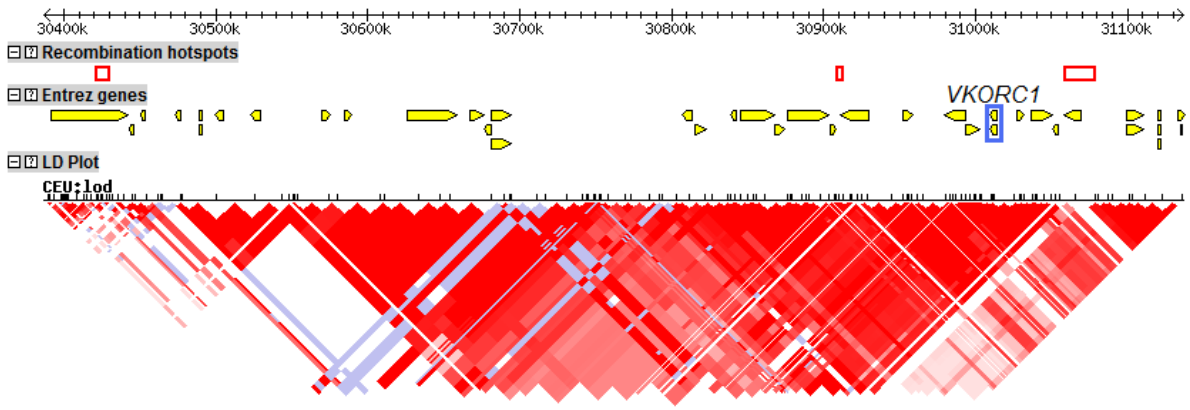


Figure 3.10. LD map of *VKORC1* in Hapmap CEU panel. Genomic architecture in Chr16:30386726 to 31136725 whereas *VKORC1* (Chr16, positions 31,009,676 to 31,013,776) is located in one end of a large LD block.

3.5.3 EPHX1

The microsomal epoxide hydrolase (EPHX1) catalyses the initial epoxide hydration reaction in the vitamin K recycling pathway (Cain et al. 1998) and forms a complex with VKOR. It is encoded by *EPHX1* on chromosome 1 (Table 3.9) which, according to Daly and King (2003), harbours functional polymorphisms that may affect protein stability and could contribute to warfarin response (Daly and King 2003). A recent study in an Israeli population has shown an association between high doses of warfarin and a coding *EPHX1* polymorphism (rs1051740) in *CYP2C9**1 wild type patients (Loebstein et al. 2005).

As shown in Table 3.14, of the 25 SNPs that passed study criteria in *EPHX1*, two are nsSNPs with rs1051740 being common in our Swedish sample. The LD map is drawn with the main contribution from 20 intronic SNPs (Table 3.14 and Figure 3.11). The genetic architecture of this locus indicates two distinct LD blocks and the presence of a recombination hot spot (Figure 3.11A), which is also seen in the Hapmap CEU panel (Figure 3.11B).

Table 3.14. Genotyping summary of *EPHX1*.

Genes	Study SNPs (tested, passed)	SNP aliases
EPHX1	42 tested, 25 passed study criteria	rs1051740 = coding 612 T>C, Y113H, increased warfarin dose requirement?; rs2234922 = coding 691 A>G, H139R, unknown function
Coding ns	2: rs1051740 (Y113H), rs2234922 (H139R)	
Coding s	2 : rs2292566, rs1051741	
UTR	0	
Intronic	18: rs2854461, rs2854447, rs2854450, rs2854451, rs3753658, rs3753659, rs3753660, rs3753661, rs2671272, rs3738047, rs2671270, rs3817268, rs2260863, rs2740170, rs4149223, rs2292567, rs2671266	
Flanking	3: rs4653436, rs3753663, rs2102663, rs6426089	

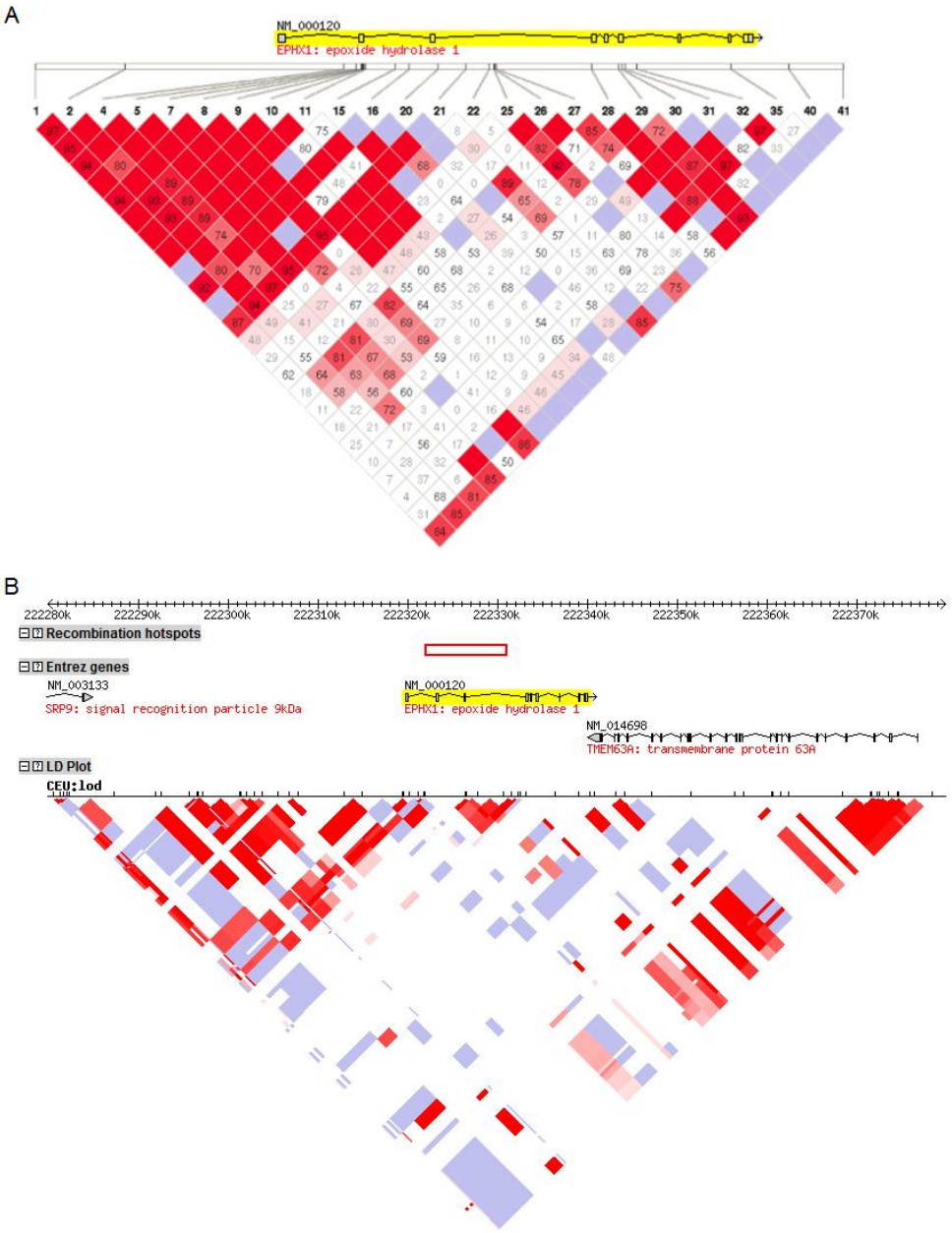


Figure 3.11. LD plots for *EPHX1* region in (A) Uppsala study and (B) Hapmap CEU panel.

3.5.4 NQO1

A few studies have shown that the antioxidant enzyme nicotine adenine dinucleotide phosphate dehydrogenase (NAD(P)H), which is also called flavoprotein DT-diaphorase (NQO1, NMOR1), reduces the quinine form of vitamin K and is of minor importance under physiologic conditions (Wallin et al. 1978). It may play a minor role in the warfarin cycle, but it was included in our study of warfarin dose.

Table 3.15 shows the genotyping summary of *NQO1*. The intronic SNP rs1437135 has been suggested to be associated with protein C level (Peyvandi et al. 2004), However, this evidence is not very strong and requires further validation. SNP rs1800566, a C-to-T substitution at position 609 of NQO1 cDNA, codes for a proline-to-serine change at residue 187 which is associated with absence of activity (Traver et al. 1992; Traver et al. 1997). This SNP is polymorphic in the Swedish population (MAF is 0.172). All SNPs passing study criteria were in LD which is consistent with the CEU Hapmap data (Figure 3.12).

Table 3.15. Genotyping summary of *NQO1*.

Genes	Study SNPs (tested, passed)	SNP aliases
NQO1	18 tested, 9 passed study criteria	rs1437135 = hCV2091258,
Coding ns	1: rs1800566 (P187S)	intronic 2515 C>T, associated
Coding s	1: rs689453	with protein C levels
UTR	0	
Intronic	6: rs2917669, rs2917671, rs1437135, rs689452, rs2965753, rs7186002	
Flanking	1: rs689456	

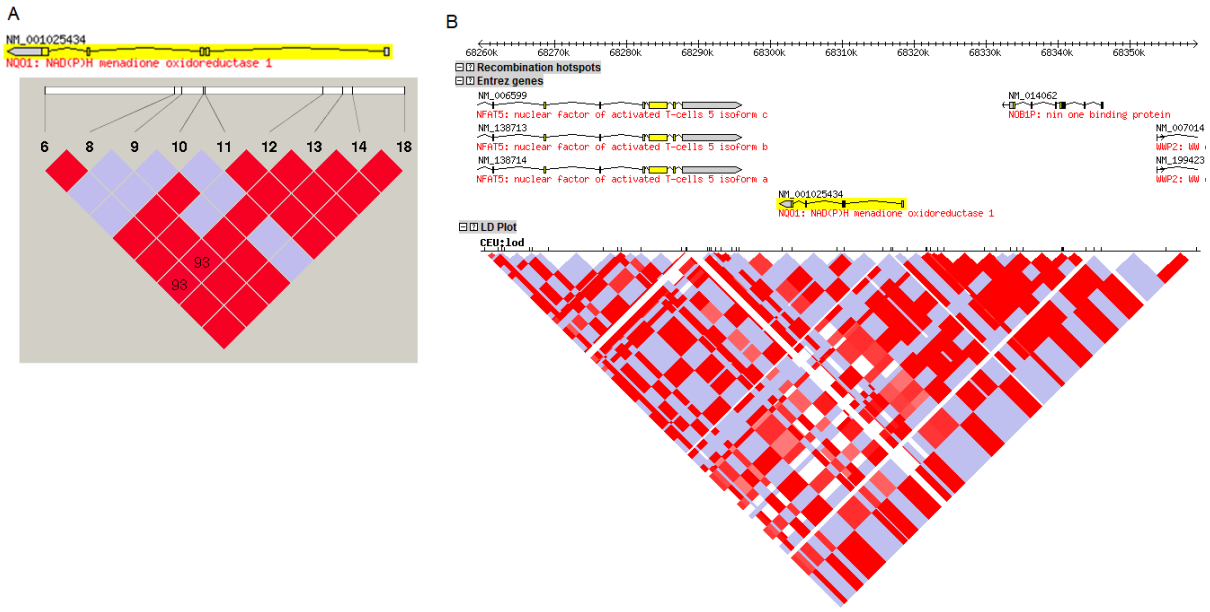


Figure 3.12. Genetic architecture of the *NQO1* locus in (A) Swedish and (B) Hapmap Caucasians. All the SNPs passing the study criteria are in LD (panel A) which is also observed in Hapmap (panel B). *NQO1* is located in the centre of a 400 kb LD block.

3.5.5 GGCX

The dietary vitamin K is an essential cofactor for the activation of coagulation proteins; which are classified as vitamin K-dependent (VKD) proteins and mediated by gamma-glutamyl carboxylase (GGCX) (Rost et al. 2004b; Wu et al. 1997) encoded by *GGCX* on chromosome 2 (Table 3.16). Gamma-glutamyl carboxylase is located in the ER membrane, spatially close to vitamin K epoxide reductase, and forms a complex with VKOR and calumenin (CALU) in order to efficiently recycle the oxidative form of vitamin K (Wajih et al. 2004).

A very rare autosomal recessive bleeding disorder has been reported with combined deficiency of the vitamin K-dependent coagulation factors II, VII, XI, and X, and proteins C, S, and Z resulting from the identification of mutations in *GGCX* (Brenner et al. 1998; Rost et

al. 2004b). The nsSNP rs699664 (C/T) which results in an arginine to glutamine substitution, has an, as yet, unknown functional impact (Table 3.16). Interestingly, the C allele is minor in Africans (MAF = ~0.30) but is major in Caucasians (MAF = 0.58) and Asians (MAF = 0.68) suggesting potential selection. Since warfarin dose requirement differs among different ethnic population, rs699664 may be of interest due to its MAF variation in different population. Table 3.16 shows the SNPs as well as a CAA microsatellite repeats that passed study criteria in *GGCX*.

Table 3.16. Genotyping summary of *GGCX*.

Genes	Study SNPs (tested, passed)	SNP aliases
GGCX	17 tested, 10 passed study criteria	rs699664 (C/T) = coding 8762
Coding ns	1: rs699664 (R325Q)	G>A (formerly 8016 G>A) or 1002
Coding s	1: rs2592551	G>A, R325Q, function unknown
UTR	0	
Intronic	6: rs7568458, rs12714145, rs6738645, rs762684, rs2028898, (CAA) microsatellite	
Flanking	2: rs6547621, rs7605975	

Six alleles of 10, 11, 13, 14, 15 and 16 CAA repeats were identified in *GGCX* intron 6 whereas only 10-, 11- and 13-repeats alleles were reported in Japanese (Shikata et al. 2004). All SNPs in *GGCX* are in LD. For illustration purposes the genomic structure of *GGCX* and LD architecture of this locus in the Swedish population is discussed alongside the results of the association analysis in chapter 4, section 4.3.3. Hapmap results show that *GGCX* is located in the centre of a 70 kb LD block.

Figure 3.13. Genomic architecture of GGCX region in Hapmap CEU panel. GGCX is located in the centre of a 70 kb LD block.

3.5.6 CALU

Calumenin is an ER chaperone protein, encoded by *CALU* on chromosome 7 (Table 3.11), which is associated with gamma-carboxylase. Recent studies have shown that calumenin binds to both VKOR and gamma-carboxylase and acts as an inhibitory protein (Wajih et al. 2004; Wallin et al. 2001). An RNAi experiment showed that the enzymatic activity of gamma-carboxylase increases five times by interfering with CALU mRNA transcription (Wajih et al. 2004). Experiments in rats indicated that over-expression of calumenin prevents

warfarin from interacting with VKOR (Wallin et al. 2001). This protects the rat from rodenticide made of high dose of warfarin. It is suggested that abundant calumenin associated to VKOR would change the conformation and prevent the affinity of warfarin targeting. However, the molecular mechanism of the warfarin resistance is still unclear.

A non-synonymous SNP (rs2290228) in exon 2 has been reported to be associated with warfarin dose (Vecsler et al. 2006), which is also polymorphic in our Swedish patients with an MAF = 0.144 (Table 3.17).

Table 3.17. Genotyping summary of *CALU*.

Genes	Study SNPs (tested, passed)	SNP aliases
CALU	14 tested, 9 passed study criteria	rs2290228 = coding 11 G>A, R4Q;
Coding ns	2: rs2290228 (R4Q), rs2307040 (A82V)	
Coding s	0	
UTR	2: rs11653, rs8597	
Intronic	4: rs2060717, rs339054, rs1006023, rs339098	
Flanking	1: rs339057	

The *CALU* gene is spanned by a single LD block in the Swedish population and Hapmap Caucasians (Figure 3.14). This LD block also contains the gene encoding opsin 1 (*OPN1SW*).

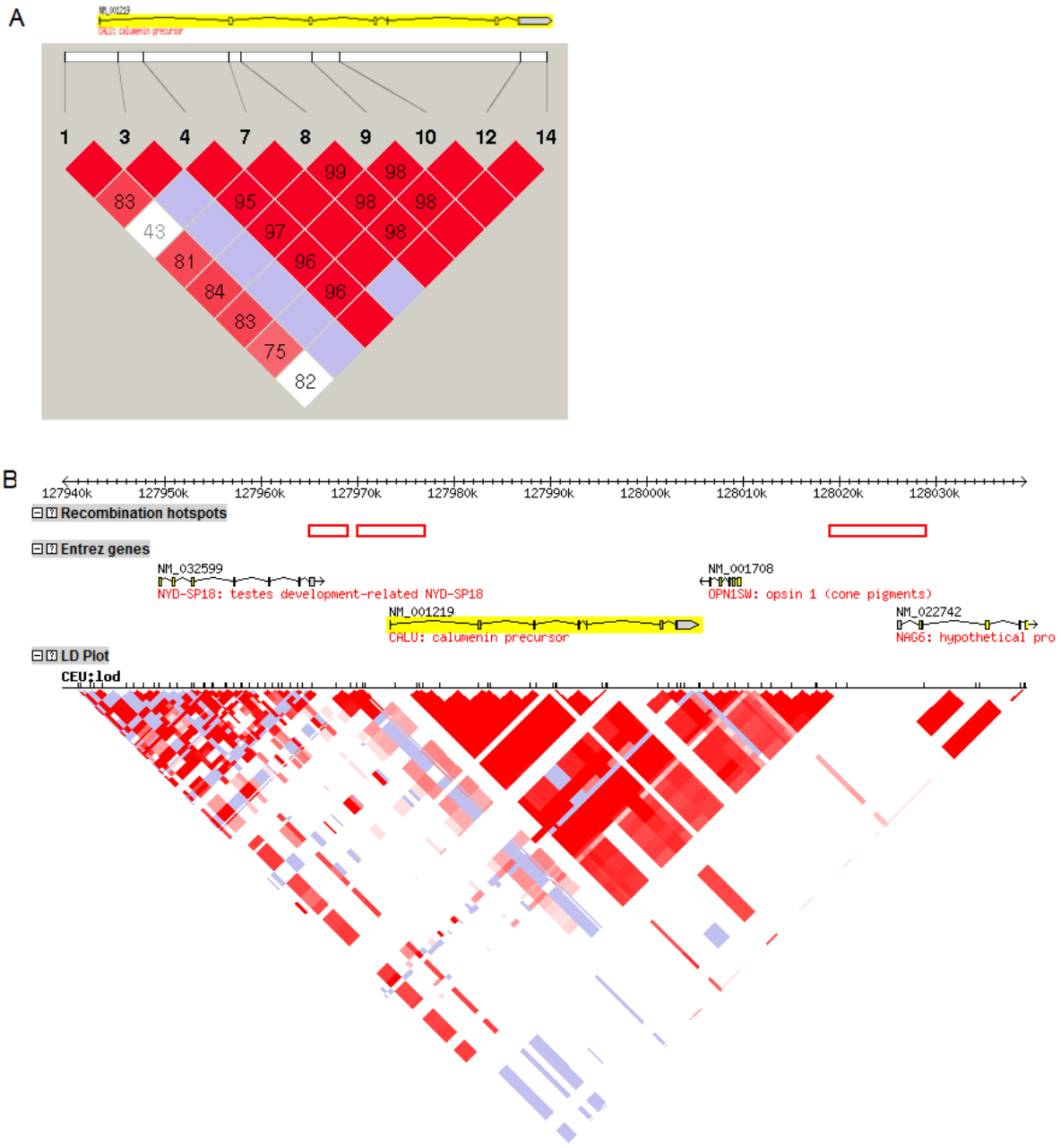


Figure 3.14. Genomic architecture of *CALU* in (A) Swedish and (B) Hapmap CEU. The Swedish LD indicates the SNPs that passed the study criteria are in LD. *CALU* and another gene, *OPN1SW*, are located in a small LD block whereas *CALU* spans across two third of this region.

3.5.7 Protein Disulfide Isomerase (PDI)

Although VKORC1 has been identified as a target for warfarin (Li et al. 2004; Rost et al.

2004a) and the involvement of its thioredoxin-like CXXC centre in the reduction of vitamin K1 2,3-epoxide (Vit.K>O) has also been validated (Wajih et al. 2005), the cellular system providing electrons to the centre was unknown. In 2007, a study using *in vitro* inhibition assay and immunoprecipitation experiment showed that a protein disulfide isomerase (PDI) peptide is tightly associated with VKORC1 and form VKOR complex (Wajih et al. 2007). Wajih and his colleagues further concluded that the energy required for gamma-carboxylation of proteins is provided by dithiol-dependent oxidative protein folding by PDI in the ER. Their proposed model present the PDI as another subunit of the complex in the ER which provides the electrons for the reduction of the thioredoxin-like CXXC centre in VKORC1 (Figure 3.15).

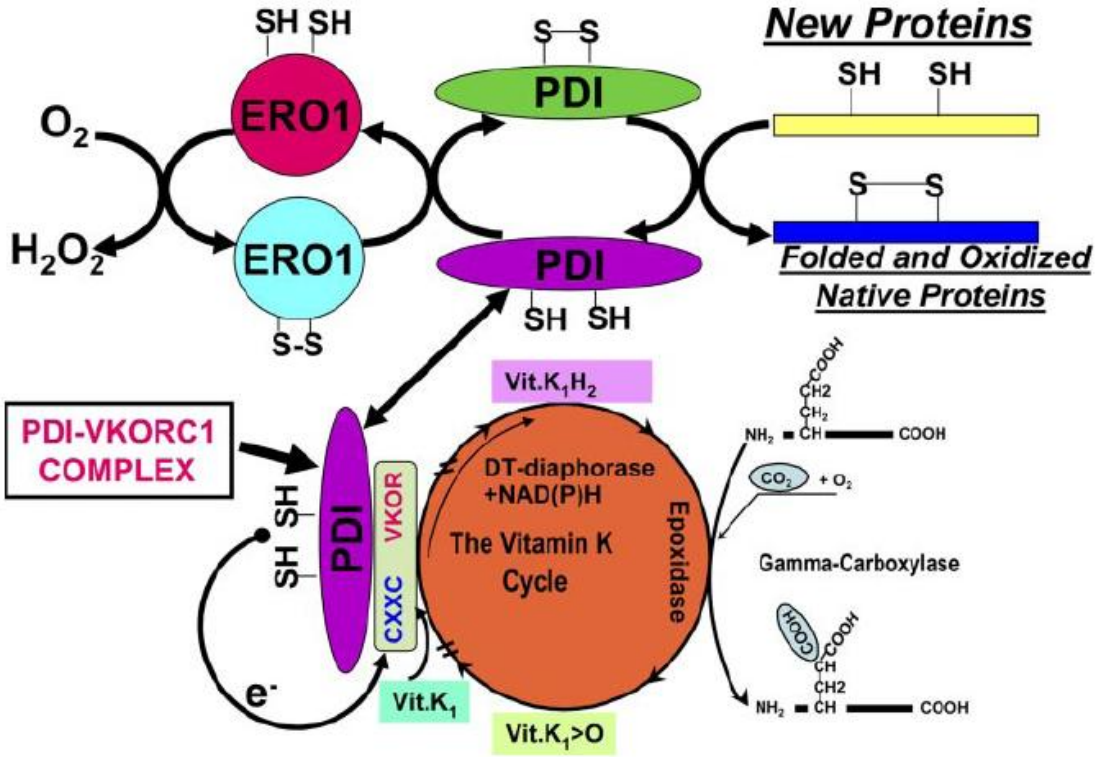


Figure 3.15. Hypothetical model of protein disulfide isomerase functionality. This figure is reproduced from J. Biol. Chem., 2007, 282(4): 2626-2635 (Wajih et al. 2007)

According to the experimental results, it was also proposed that part of the PDI peptide pool in the ER becomes strongly linked to part of the VKORC1 pool and that the complex is the active warfarin-sensitive Vit.K>O reducing enzyme complex of the vitamin K cycle. Since PDI is found in abundance in all eukaryotes and is fairly uniform in its main features, this proposed model also explains the VKOR activity is inhibitory in transfection expression study in various cell lines including insect cells (Li et al. 2004).

In 2006, a study suggested that oxidation/isomerisation by PDI switches tissue factor from a non-coagulant to an active molecule (Ahamed et al. 2006). Later, Versteeg and Ruf reported that tissue factor (TF) mediated coagulation is significantly enhanced and factor X activation is 5-10 times increased with presence of PDI (Versteeg and Ruf 2007). Although Wajih et al (Wajih et al. 2007) described the use of siRNA to inhibit disulfide isomerase activity, no precise DNA or protein sequence was reported. In Homo *sapiens*, there are six genes known as protein disulfide isomerases. In the absence of any further information all known *PDI* genes were tested for association to warfarin-induced bleeding and, in parallel, warfarin dose requirement. Table 3.18 lists a total of 31 SNPs that passed study criteria. The six PDI genes were selected in a later phase of the project. SNPs genotyped in Hapmap were chosen, except for *P4HB* where only one SNP was available with a further two SNPs within 200 kb flanking region. SNPs in *P4HB* were selected from the Ensembl SNP database. Hapmap CEU result indicates that *PDIA5* spans across three LD blocks, i.e. two recombination hot spots were observed in the gene. However, due to selection of tag SNPs, our result in the Swedish population is not reflecting the LD structure properly (Figure 3.17). LD structures of the six PDIs in the Swedish sample are shown in Figure 3.17.

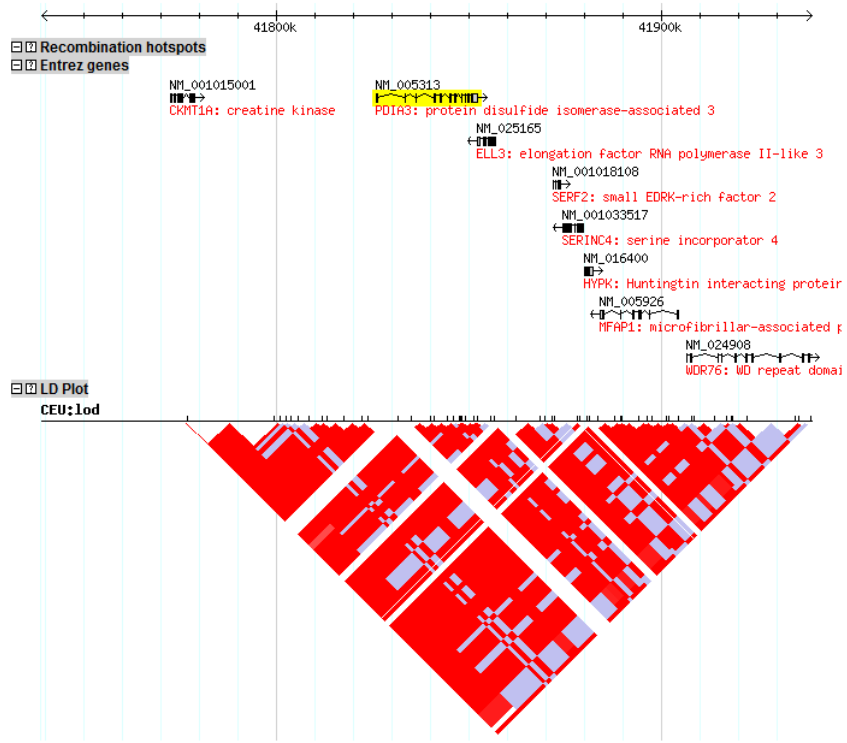
Table 3.18. Genotyping summary of *PDI* family.

Genes	Study SNPs (tested, passed)	SNP aliases
P4HB	6 tested, 5 passed study criteria	-
Coding ns	0	
Coding s	2: rs1130674, rs2070871	
UTR	0	
Intronic	3:rs876017, rs1533756, rs1010954	
Flanking	0	
PDIA2	4 tested, 2 passed study criteria	rs2685127, function unknown; rs400037 (R388Q), function unknown
Coding ns	2: rs2685127 (T286M), rs400037 (R388Q)	
Coding s	0	
UTR	0	
Intronic	0	
Flanking	0	
PDIA3	4 tested, 4 passed study criteria	-
Coding ns	0	
Coding s	0	
UTR	0	
Intronic	4: rs10163054, rs8040336, rs11070411, rs7175032	
Flanking	0	
PDIA4	8 tested, 7 passed study criteria	-
Coding ns	0	
Coding s	0	
UTR	0	
Intronic	7: rs10085877, rs4727005, rs10272564, rs10269104, rs6464929, rs1551927, rs6464930	
Flanking	0	
PDIA5	6 tested, 6 passed study criteria	-
Coding ns	0	
Coding s	0	
UTR	0	
Intronic	6: rs1078982, rs3792366, rs4677875, rs702030, rs836832, rs1107377	
Flanking	0	
PDIA6	7 tested, 7 passed study criteria	-
Coding ns	0	
Coding s	0	
UTR	0	
Intronic	6: rs1198873, rs11904084, rs1686447, rs1734343, rs1734346, rs12471762	
Flanking	1: rs1686482	

A


180k 190k 200k 210k 220k 230k 240k 250k 260k 270k 280k 290k 300k 310k 320k 330k 340k 350k 360k 370k
Recombination hotspots
Entrez genes
NM_201412
LUC7L: LUC7-like
NM_183337
RGS11: regulator of G-protein signalling 11
NM_032039
C16orf9: chromosome 16 open reading frame 9
NM_006428
MRPL28: mitoch
NM_001176
ARHGDIG: Rho GDP dissociation inhibitor (GDI) gamma
NM_006849
PDIA2: protein disulfide isomerase-associated 2
NM_181050
AXIN1: axin 1
LD Plot
CEU:lod

B


41800k
41900k
Recombination hotspots
Entrez genes
NM_001015001
CKMT1A: creatine kinase
NM_005313
PDIA3: protein disulfide isomerase-associated 3
NM_025165
ELL3: elongation factor RNA polymerase II-like 3
NM_001018108
SERF2: small EDRK-rich factor 2
NM_001033517
SERINC4: serine incorporator 4
NM_016400
HYPK: Huntingtin interacting proteir
NM_005926
MFAP1: microfibrillar-associated p
NM_024908
WDR76: WD repeat domai
LD Plot
CEU:lod

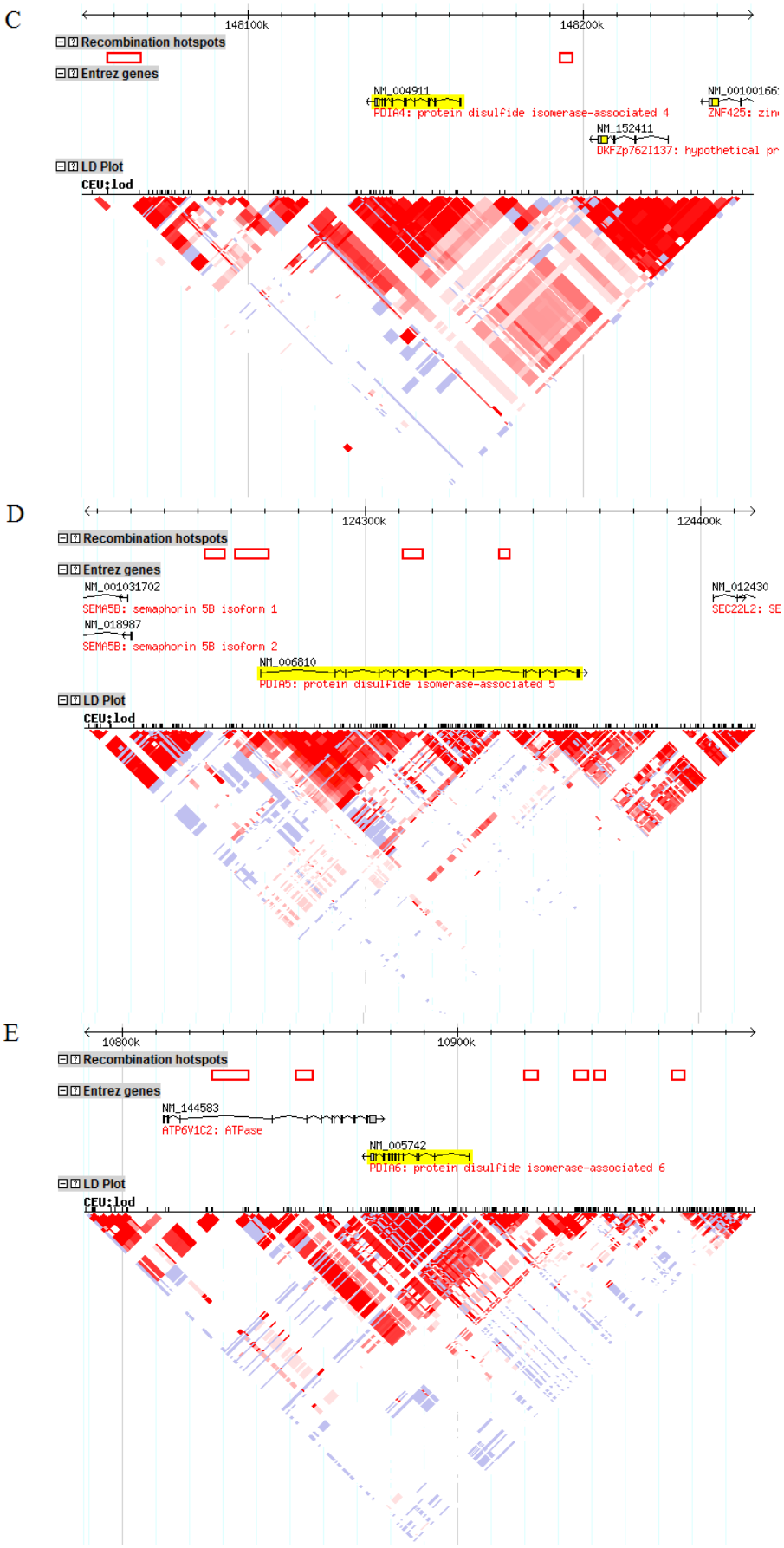


Figure 3.16. Genomic architectures of (A) *PDIA2*, (B)*PDIA3*, (C) *PDIA4*, (D) *PDIA5* and (E) *PDIA6* in Hapmap. Within 200 kb flanking near *P4HB*, only two SNPs were genotyped in Hapmap.

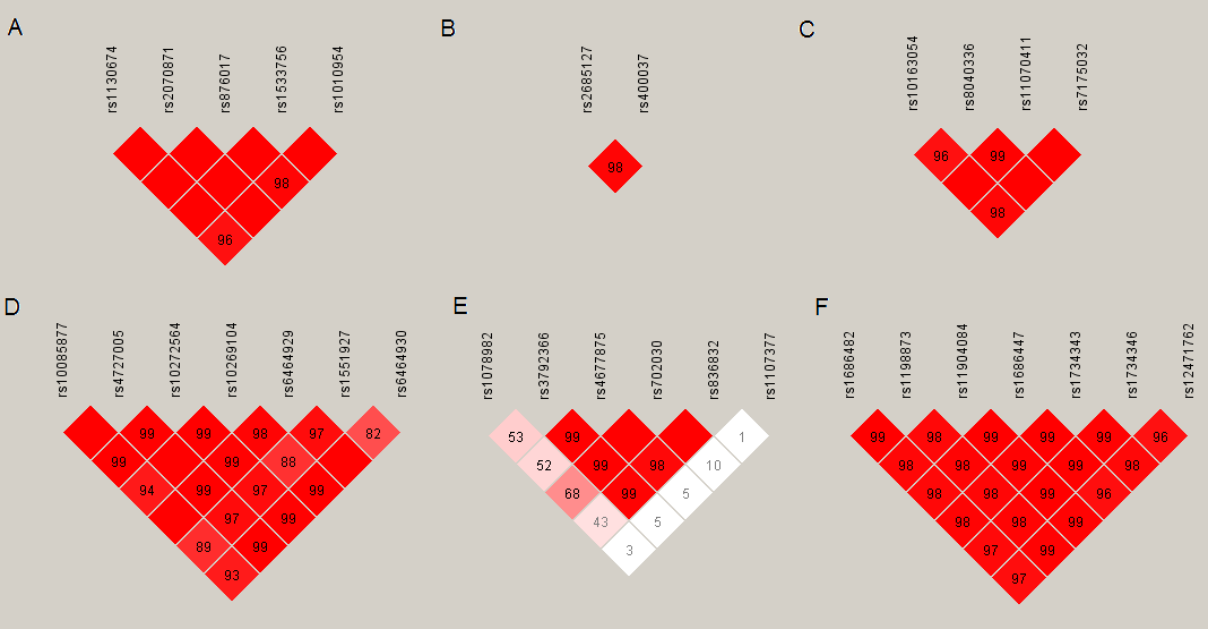


Figure 3.17. LD structures of six genes in PDI family A. (A) *P4HB*, (B) *PDIA2*, (C) *PDIA3*, (D) *PDIA4*, (E) *PDIA5* and (F) *PDIA6*.

3.6 Vitamin K Dependent Proteins

The role of many vitamin K-dependent (VKD) proteins has been investigated in warfarin sensitivity and, at the start of this study, several had been postulated to be implicated (see below). The main VKD proteins are clotting factors II (prothrombin), VII, IX and X, proteins C, S and Z and growth-arrest-specific protein 6, encoded by *F2, F7, F9, F10, PROC, PROS1, PROZ and GAS6* (Berkner 2000; Berkner and Runge 2004). Table 3.19 shows the genomic information for each candidate VKD gene. *F7*, *F10*, and *PROZ* are located back to back in chromosome 13:112808124-112874700 whilst *GAS6* is located 672 kb telomeric to this cluster.

Table 3.19. Candidate genes of vitamin K dependent.

Gene	Location	Exons	Transcript	Protein (residues)	Function of protein
F2	Chr 11: 46697331 - 46717631 bp	14	1997 bps	622	Converts fibrinogen to fibrin, activates FV, FVIII, FXIII, protein C
F7	Chr 13: 112808124 - 112822348 bp	9	2459 bps	466	Is converted to FVIIa and then converts FIX to FIXa and FX to FXa
F9	Chr X: 138340437 - 138373137 bp	8	2780 bps	461	Makes a complex with FVIIIa and then converts FX to its active form
F10	Chr 13: 112825128 - 112851846 bp	8	1524 bps	488	Converts FII to FIIa in the presence of factor Va.
PROC	Chr 2: 127892246 - 127903048 bp	9	1756 bps	461	Activated protein C counteracts coagulation together with protein S by inactivating FVa and VIIIa
PROS1	Chr 3: 95074647-95175395 bp	15	3275 bps	676	Cofactor to protein C that degrades coagulation factors Va and VIIIa
PROZ	Chr 13: 112860971 - 112874700 bp	8	1488 bps	400	Is together with protein Z-dependent protease inhibitor a cofactor for the inactivation of FXa
GAS6	Chr 13: 113546903 - 113590421	15	2499 bps	678	Participates in many processes, e.g. potentiation of agonist-induced platelet aggregation

3.6.1 Gene cluster on chromosome 13 (F7, F10, PROZ)

Factor VII also known as Hageman factor, is a serine protease encoded by the *F7* gene on chromosome 13. In both intrinsic and extrinsic pathways, factor VII is the first to be activated in trauma (extrinsic pathway) and initiate the coagulation cascade (Figure 3.18). A few studies have reported promoter polymorphisms in *F7* to have an effect on warfarin sensitivity (Aquilante et al. 2006; D'Ambrosio et al. 2004; Shikata et al. 2004).

Factor X is also known as Stuart-Prower factor or as thrombokinase and is equipped with serine endopeptidase activity. It is encoded by the *F10* gene located adjacent to *F7*. Studies of promoter variants and a synonymous coding polymorphism in exon 7 of *F10* have reported no effect on warfarin sensitivity (Aquilante et al. 2006; Shikata et al. 2004). However, factor X plays an important role as the first member of the final common pathway (also called thrombin pathway) (Figure 3.18).

Protein Z, encoded by *PROZ*, is a glycoprotein and assists hemostasis by binding thrombin and promoting its association with phospholipid vesicles. It also accelerates 1000-fold of the degradation of factor X which is primarily done by protein Z-related protease inhibitor (ZPI). Mutation in ZPI has been reported to be associated with an increased susceptibility to venous embolisms (Van de Water et al. 2004).

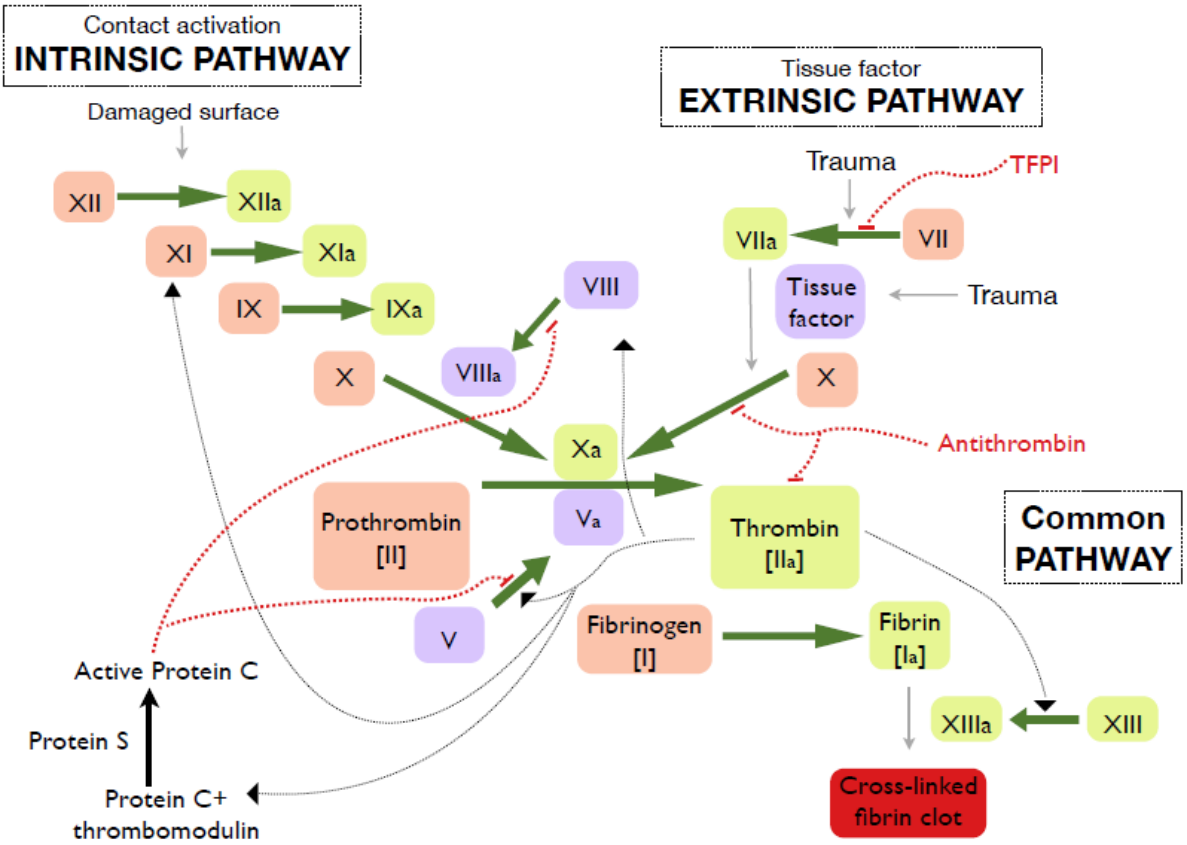


Figure 3.18. The coagulation cascade. (This figure also appears in chapter 1 as Figure 1.3, modified from http://en.wikipedia.org/wiki/Coagulation).

As described previously, *F7*, *F10*, and *PROZ* are sitting back to back on chromosome 13. In this 80 kb region genotype results were obtained for 39 SNPs (Table 3.20). An nsSNP in *F7* (rs6046, R413Q) has been previously suggested to give rise to a peptide with lower activity in blood coagulation. LD analysis shows that *F7* and the first exon of *F10* are located in the same LD block, whereas the last two exons of *F10* sit in the same LD block with *PROZ* (Figure 3.19A). It is evident that there is a recombination hot spot near the intron 6 of *F10* whereas exons 2, 3, 4 and 5 of *F10* appear to form a sub haplotype block suggesting the presence of a weak recombination hot spot.

The result from Hapmap CEU panel replicates our finding in 201 Swedish patients (Figure 3.19B). Two recombination hot spots were observed in *F10*. The Hapmap result further suggests a genomic instability in the region including *F7* and *F10* whereas *PROZ* is located in a distinct LD block (Figure 3.19B).

The Online Mendelian Inheritance in Man (OMIM) database reports 24 different mutations in *F7* of which 22 are associated with deficiency or decrease enzymatic activity and two with a decreased susceptibility to myocardial infarction. In *F10*, 14 mutations are reported to result in decreased activity or deficiency and which cause mild or severe bleeding. In contrast to *F7* and *F10*, *PROZ* has not been reported to harbour any variants associated with disease state.

Table 3.20. Genotyping summary of *F7*, *F10* and *PROZ* cluster.

Genes	Study SNPs (tested, passed)	SNP aliases
F7	32 tested, 11 passed study criteria	rs6046 = coding 1289 G>A, R413Q (formerly 1238 G>A, R353Q), low activity allele
Coding ns	1: rs6046 (R413Q)	
Coding s	0	
UTR	1: rs2476324	
Intronic	6: rs2774030, rs491098, rs493833, rs569557, rs488703, rs6041	
Flanking	3: rs3093229, rs3093230, rs3093233	
F10	45 tested, 15 passed study criteria	rs5960 = synonymous coding T264T, function unknown
Coding ns	0	
Coding s	1: rs5960	
UTR	0	
Intronic	12: rs473598, rs776897, rs3211770, rs2026160, rs2251102, rs3211764, rs2480946, rs693335, rs483949, rs485798, rs776905, rs474810	
Flanking	2: rs563964, rs3093261	
PROZ	27 tested, 13 passed study criteria	rs3024711 = rs17878660, intronic
Coding ns	0	
Coding s	0	
UTR	0	
Intronic	11: rs3024764, rs3024747, rs3024746, rs17881956, rs3024743, rs17886440, rs3024731, rs2480948, rs513479, rs3024718, rs3024711	
Flanking	2: rs2273971, rs7335409	

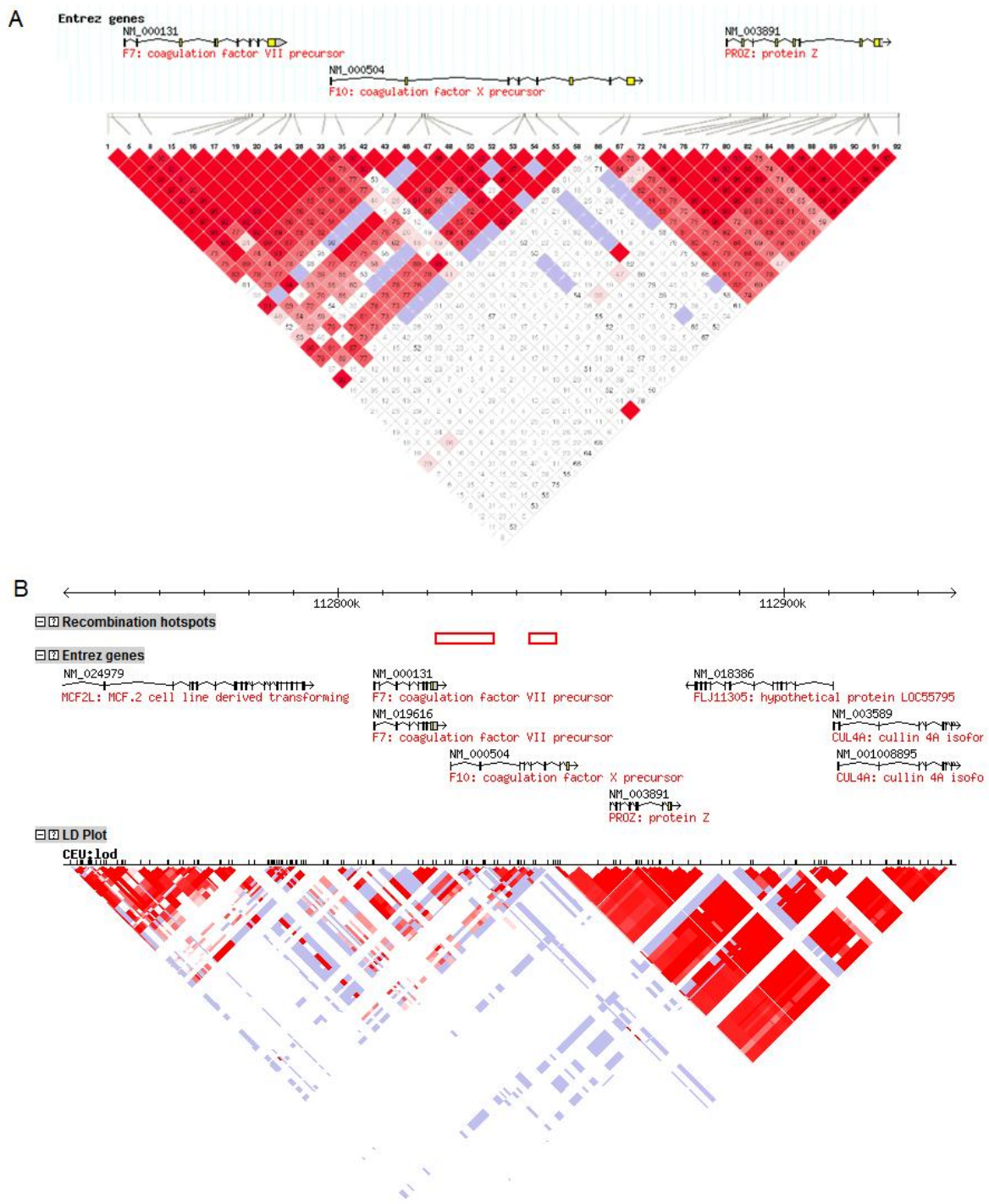


Figure 3.19. Genomic architecture and LD organisation of the chromosome 13 locus harbouring *F7*, *F10* and *PROZ* in (A) Swedish and (B) Hapmap Caucasians.

3.6.2 Other VKD genes (F2, F9, PROC, PROS1 and GAS6)

Beside the chromosome 13 cluster, the following VKD proteins were selected and genotyped as candidate genes: coagulation factor II and IX which are involved in the coagulation cascade; protein C and Z which act as natural anticoagulants by inhibiting factor V and VIII, and finally growth arrest-specific gene 6 which activates platelet formation (Table 3.21).

Table 3.21. Genotyping summary of other VKD candidate genes.

Genes	Study SNPs (tested, passed)	SNP aliases
F2	22 tested, 10 passed study criteria	rs5896 = coding 525 C>T (formerly 494 C>T), T165M, warfarin sensitivity?
Coding ns	1: rs5896 (T165M)	
Coding s	1: rs5898	
UTR	0	
Intronic	8: rs2070850, rs3136435, rs3136447, rs2070851, rs2070852, rs3136460, rs2282687, rs3136516	
Flanking	0	
F9	29 tested, 11 passed study criteria	Chr X: 138344709: coding G>A, A-10T; Chr X: 138344710: coding C>T, A-10V; Both are rare causes of warfarin sensitivity; These mutations were not found in the study
Coding ns	1: rs6048 (T194A)	
Coding s	0	
UTR	1: rs440051	
Intronic	7: rs401597, rs392959, rs398101, rs422187, rs413957, rs110583, rs413536	
Flanking	1: rs434447, rs445691	
PROC	25 tested, 13 passed study criteria	rs1799809 = upstream -1644 A>G (formerly -1641), GG lower protein C activity
Coding ns	0	
Coding s	1: rs5936	
UTR	0	
Intronic	10: rs2069910, rs2069915, rs2069916, rs2069919, rs2069921, rs973760, rs1518759, rs2069924, rs2069928, rs2069931	
Flanking	2: rs2069901, rs1799809	
PROS1	50 tested, 11 passed study criteria	-
Coding ns	0	
Coding s	0	
UTR	0	
Intronic	8: rs8178592, rs5013930, rs8178607, rs8178610, rs4857343, rs8178623, rs4857037, rs8178649	
Flanking	3: rs7650230, rs9713061, rs9683303	
GAS6	5 tested, 4 passed study criteria	-
Coding ns	0	
Coding s	0	
UTR	0	
Intronic	3: rs9577874, rs9604573, rs6602908	
Flanking	1: rs7997328	

1. Coagulation

Prothrombin (factor II) is encoded by *F2*. The activated form (thrombin, factor IIa) converts soluble fibrinogen into insoluble fibrin and activates and further amplifies coagulation cascade by activating factor XI, VIII, and V (Figure 3.18). Over-presence of thrombin will also activate protein C for anticoagulation. Two independent studies have shown that a polymorphism in F2 causing a change from threonine to methionine at residue 165 leads to increased sensitivity to warfarin (D'Ambrosio et al. 2004; Shikata et al. 2004) whereas a third study could not replicate this finding (Aquilante et al. 2006).

Factor IX is also a serine protease and activates factor X in the intrinsic pathway. To date, 110 mutations found in the *F9* gene, which is located on the long arm of the X chromosome, have been registered in the OMIM database. Factor IX deficiency causes Hemophilia B which occurs in the general population at one-sixth the incidence of Hemophilia A, affecting approximately 1 in 30,000 male births. Mutations in the propeptide of factor IX, amino acid substitution change from alanine to valine or threonine at residue 10, lead to a rapid drop of factor IX during warfarin treatment resulting in warfarin sensitivity and thereafter bleeding in rare cases (Kristensen 2002).

The LD result of *F2* and *F9* is shown in Figure 3.20. Our Swedish results indicated both genes are well located in LD blocks and this is confirmed by the Hapmap CEU result. Additionally, Hapmap CEU results indicate that the LD block, where F2 is located, is bigger than 1 Mb whereas F9 is located at the end of a 150 kb block.

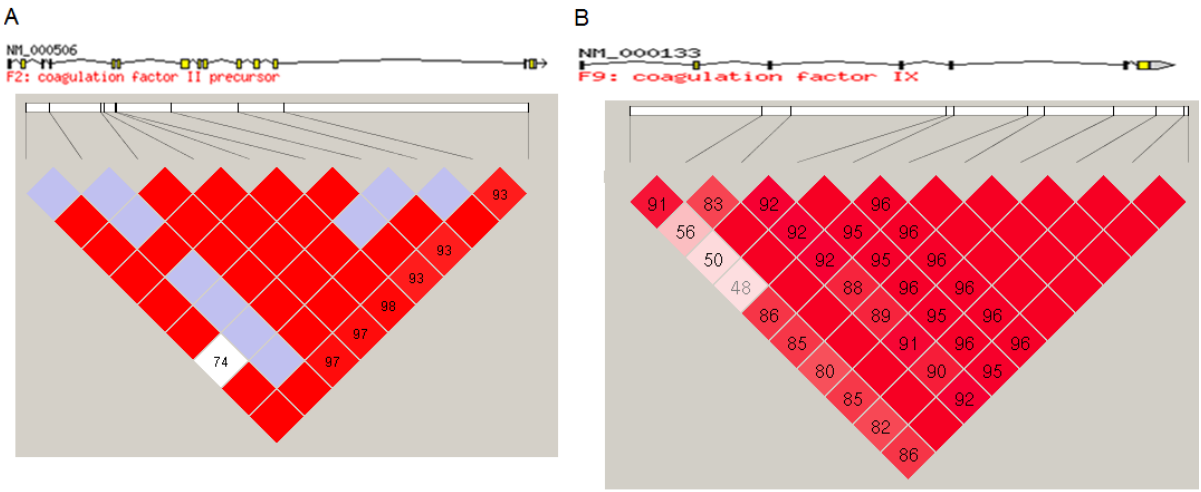


Figure 3.20. LD architecture of (A) F2 and (B) F9.

2. Anticoagulation

Protein C, encoded by *PROC* on chromosome 2, is a VKD serine protease that regulates blood coagulation by inactivating factors Va and VIIIa in the presence of calcium ions and phospholipids (Berkner 2000; Dahlback 2005). Protein S is a VKD plasma glycoprotein synthesized in the liver and is encoded by *PROS1* on chromosome 3. In anticoagulation, it functions as a cofactor to the activated protein C (APC) in the degradation of coagulation factors Va and VIIIa (Figure 3.18) and helps to prevent coagulation and stimulates fibrinolysis (Berkner 2000; Dahlback 2005).

Deficiency of protein C and S has been reported to predispose to thrombophilia (Engesser et al. 1987; Hasstedt et al. 1998; Matsuda et al. 1988). Protein C and S levels decline more rapidly than other VKD proteins resulting in a poor antithrombotic efficacy in the beginning of warfarin treatment (Vigano et al. 1984; Weiss et al. 1987). Warfarin-treated patients with a hereditary deficiency of protein C or S might result in warfarin-induced hypercoagulation (Chan et al. 2000; McGehee et al. 1984) and skin necrosis (Conlan et al. 1988). SNPs that

passed study criteria were evenly distributed alongside both genes (Figure 3.21). Both genes are in the same LD block; which is in agreement with Hapmap.

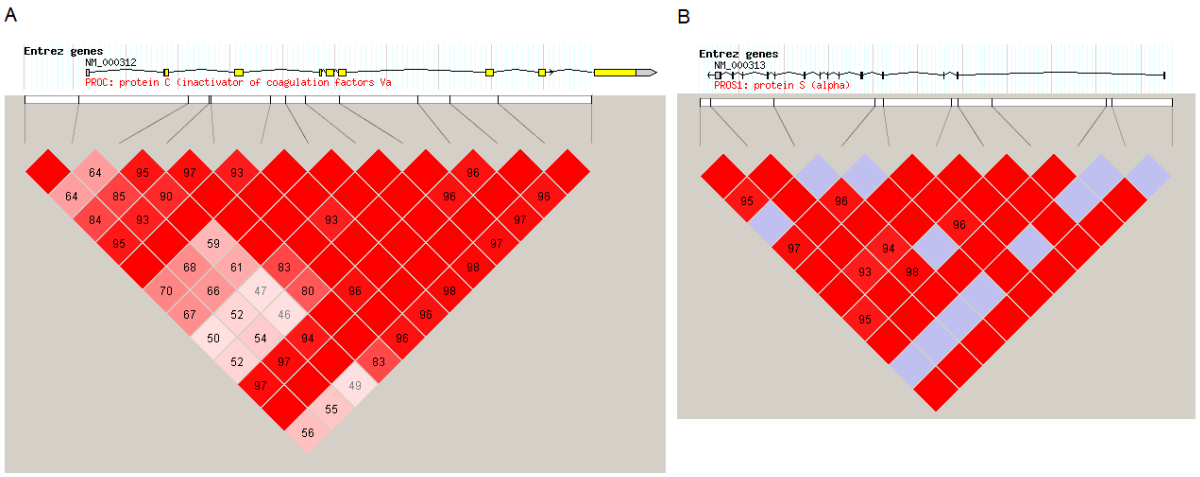


Figure 3.21. Genomic architecture of protein C and protein S in Swedish.

3. Platelet Formation

Growth arrest-specific gene 6 (*GAS6*) was first identified in 1993 with a 44% sequence similarity to coagulation protein S, and to which it is structurally related but without the anticoagulant activity (Manfioletti et al. 1993). Like protein S, GAS6 is also a vitamin K-dependent protein and is secreted from platelet A-granules and, it has been suggested to be involved in platelet activation and thrombus development (Angelillo-Scherrer et al. 2001; Gould et al. 2005).

Five SNPs in *GAS6* were reported to be significantly associated with atherothrombotic disease in a group of 110 healthy controls and 188 patients (Munoz et al. 2004). Together with being a VKD protein, *GAS6* may be associated with warfarin dose.

We obtained 4 SNPs that passed study criteria (Table 3.21) and of those 3 are in strong LD (Figure 3.22A). Hapmap CEU panel also indicate the presence of LD breaks in this 200 kb region. Three recombination hot spots are located in *GAS6* and 30 kb flanking region of both ends.

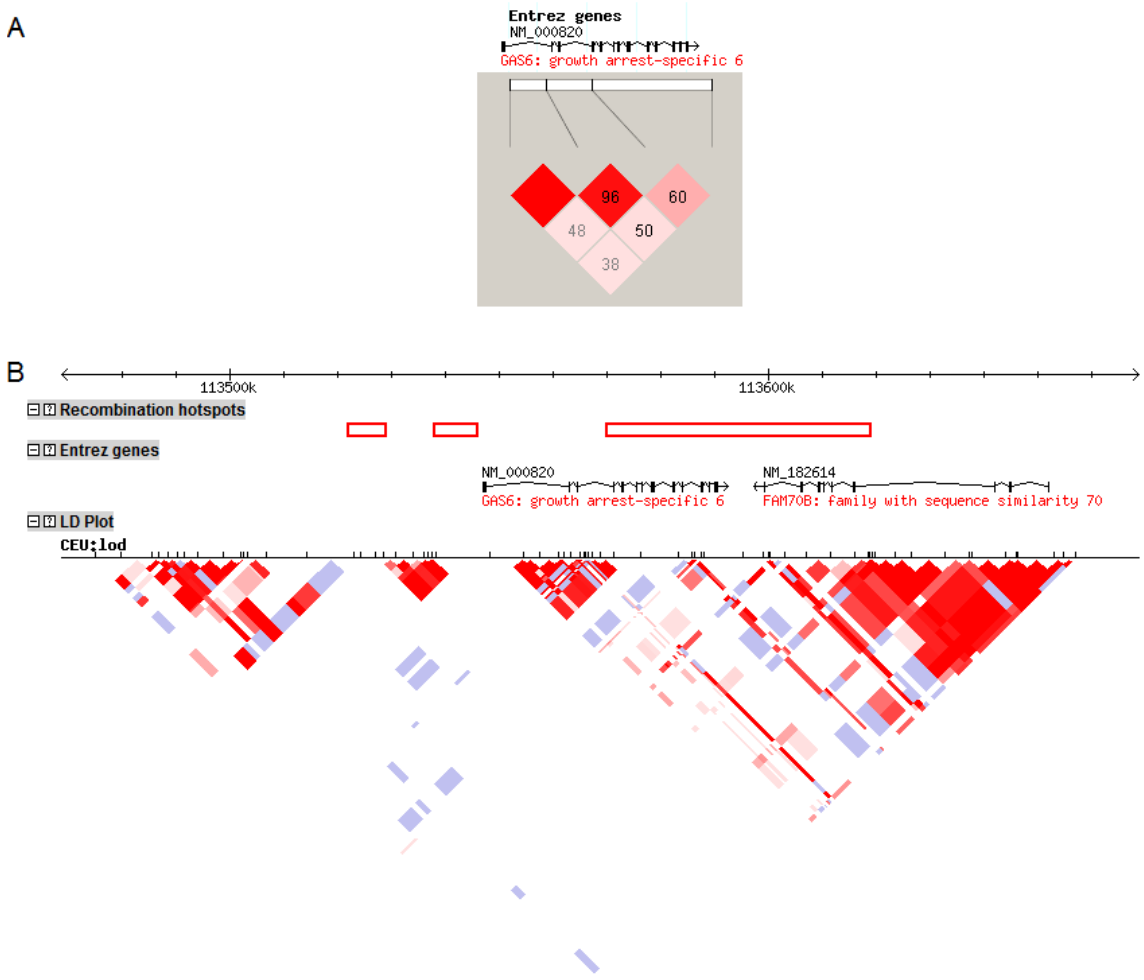


Figure 3.22. Genomic structure of growth arrest-specific gene 6 (GAS6) peptide in (A) Swedish and (B) Hapmap CEU panel.

3.7 OTHER COAGULATION FACTORS

Two other genes, which are important in the coagulation cascade, were selected (Table 3.22). The first one is antithrombin-III precursor encoded by *SERPINC1* on chromosome 1 which inhibits factors II, IX, X, XI and XII. The second one is factor V encoded by *F5* also located on chromosome 1 which activates prothrombin (factor II). A total of 97 SNPs were tested and 50 SNPs passed the study criteria (Table 3.23). Results of seven non-synonymous SNPs in factor V were obtained including FV Leiden mutation which confers increased risk of thrombosis (Larsen et al. 1998).

Table 3.22. Other candidate genes in this study.

Gene	Location	Exons	Transcript	Protein (residues)	Function of protein
SERPINC1	Chr 1: 170604596 - 170618130 bp	7	1559 bps	464	Inhibits FIIa, FIXa, Xa, XIa and XIIa. Anti-thrombin deficiency increases risk of thrombosis
F5	Chr 1: 166215067 - 166287379 bp	25	6914 bps	2224	A cofactor that activates FII together with FXa. A F5 mutation leads to risk of thrombosis

Table 3.23. Genotyping summary of antithrombin III and factor V.

Genes	Study SNPs (tested, passed)	SNP aliases
SERPINC1	23 tested, 9 passed study criteria	-
Coding ns	0	
Coding s	1: rs5878	
UTR	0	
Intronic	6: rs2227590, rs2227593, rs2227594, rs2227607, rs5877, rs2759328	
Flanking	2: rs2227588, rs2146372	
F5	74 tested, 41 passed study criteria	rs6025 = coding 1698G>A, R534Q (formerly R506Q), FV Leiden increased risk of thrombosis
Coding ns	7: rs6033 (M413T), rs6025 (R534Q), rs6018 (N817T), rs4525 (H865R), rs6032 (K925E), rs6030 (M1764V), rs6027 (D2222G)	
Coding s	9: rs6028, rs6029, rs6022, rs6035, rs6015, rs6036, rs6037, rs6024, rs6021	
UTR	0	
Intronic	23 : rs3753305, rs9332504, rs2298905, rs2298908, rs2236870, rs3766121, rs3766120, rs3766119, rs1894702, rs6012, rs3766117, rs1894699, rs6427198, rs721161, rs2298909, rs3766110, rs1557572, rs9332618, rs9332629, rs2213867, rs2213866, rs2227244, rs966751	
Flanking	2 : rs2269648, rs2187952	

3.7.1 Anti-thrombin III

Anti-thrombin III is a serine protease inhibitor in the plasma that regulates the intrinsic coagulation pathway. The inhibition reaction traps the protease including thrombin and activated form of factor IX, X, XI, and XII, by preventing the enzymatic site from specific substrates (Olson and Bjork 1994). Anti-thrombin III deficiency is a rare inheritable disease and patients normally suffer recurrent venous thrombosis and pulmonary embolism. Both the congenital form caused by mutations in *SERPINC1* and the acquired form may result in a hyper-coagulable scenario during warfarin induction (Chan et al. 2000; Dahlback 2005). Analysis of the nine SNPs that passed study criteria shows that *SERPINC1* spans a single haplotype block (Figure 3.23). Hapmap indicates that this block is more than 1 Mb in length.

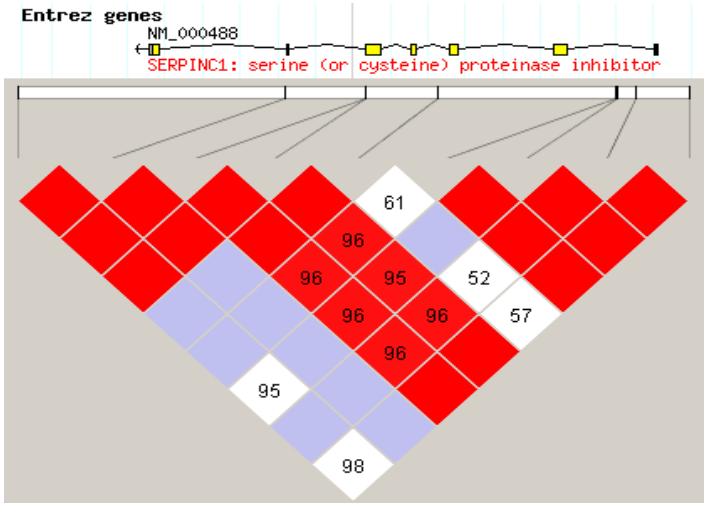


Figure 3.23. Genomic architecture of *SERPINC1*in Swedish.

3.7.2 Coagulation factor V

Factor V is a stimulatory protein without enzymatic activity which circulates in the plasma and binds to activated platelets. Factor V is activated by thrombin (FII) and the activated form of factor V reciprocally activates prothrombin to thrombin (active form) for blood

clotting. Mutations in F5 usually cause either the factor V deficiency or resistance to activated protein C. Deficiency of factor V leads to predisposition for mild or severe hemorrhage in recessive action (Guasch et al. 1998; Lindqvist et al. 1998; van Wijk et al. 2001). Meanwhile, mutations like factor V Leiden (Arg506Gln), commonly cause thromboembolism (Larsen et al. 1998) resulting a stubborn factor V protein which is resistant to cleavage by activated protein C: This keeps it constantly active and increases the rate of thrombin formation. To date, *F5* mutations are not known to be associated with warfarin treatment. *F5* also harbours an intragenic recombination hot spot (Figure 3.24).

Our result is in agreement with Hapmap CEU which also reports that LD breaks in introns 6 and 7 (Figure 3.25A). The first seven exons encode for 372 amino residues, whilst the whole peptide is 2224 residues long. This recombination hot spot is consistently observed in Han-Chinese (Figure 3.25B), Japanese (Figure 3.25C) and African from Yoruba.

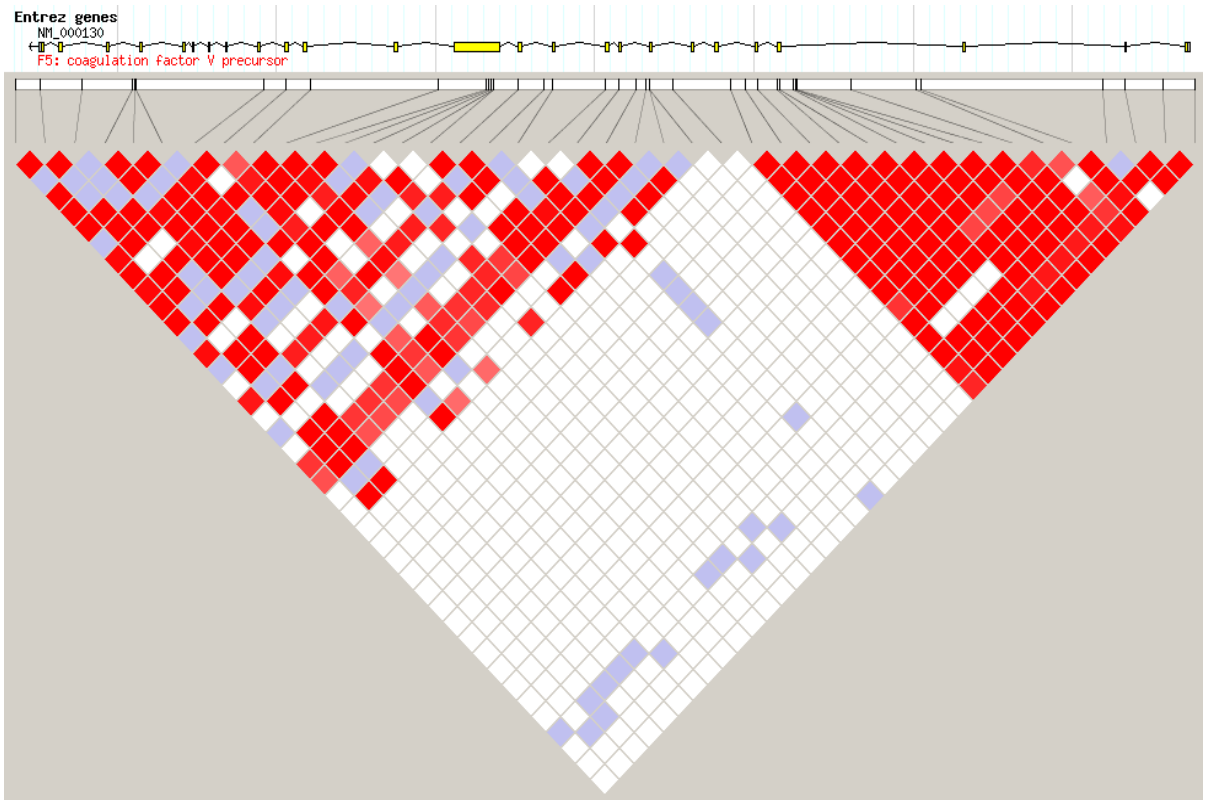


Figure 3.24. Genomic architecture of factor V in Swedish.

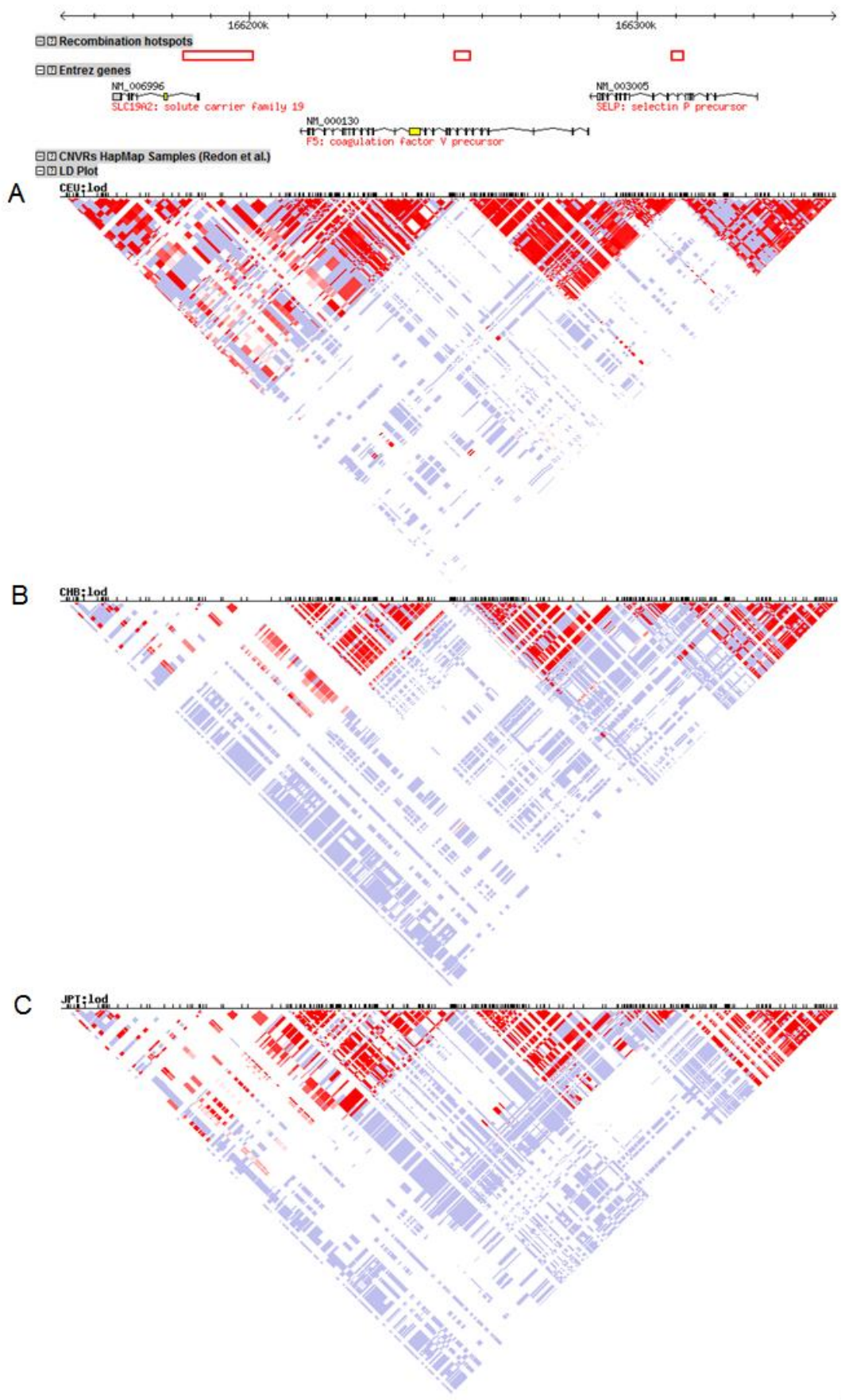


Figure 3.25. Genomic architecture of F5 region in Hapmap (A) Caucasian, (B) Han-Chinese and (C) Japanese.

Most of FV mutations have been reported in Caucasians. By comparing the Hapmap result of Caucasian (CEU), Han-Chinese (CHB) and Japanese (JPT), the LD blocks in Caucasian separate distinctly whereas marginal LD was observed between blocks in Han-Chinese and Japanese. This might suggest more severe genomic instability near the recombination hot spot within F5, which results in a higher mutation rate observed in the CEU panel.

3.8 CONCLUSION

We carefully selected 35 candidate genes that we believe have potential effect in inter-individual variation of warfarin dose and in bleeding complications of warfarin treatment. Except for genes we selected in a later phase (*APOE, GAS6* and *PDI* family) and for which we mainly used tag SNPs from Hapmap, we constructed comprehensive LD maps in our Swedish cohort(s) for each of the candidate genes. Not surprisingly, the genetic structure in the Swedish population is in accordance with that seen in the Hapmap CEU panel which reflects northern and western European ancestry.

In the first phase of this project, a lot of effort was put into validation for the SNPs deposited in public databases. In this regard, the Hapmap project provides a resource of information in genetic variation, and in later phases of the project SNPs validated by Hapmap were preferentially selected to test our patients.

All the SNPs that passed study criteria were analysed for association to warfarin dose in the 201 patients of Uppsala cohort (chapter IV). These SNPs were also used in the selection of tag SNPs to analyse warfarin-treated patients from the WARG study (chapter V), as well as the association with warfarin-induced bleeding complication (chapter VI).