The University Of Sheffield.

Investigation of the association of novel genetic factors on von Willebrand factor (VWF) plasma levels

Ahmad Hasan Mufti

Thesis Submitted for the Degree of Doctor of Philosophy

The University of Sheffield

Department of Infection, Immunity and Cardiovascular Disease

Haemostasis Research Group

January 2016

Acknowledgments

In the name of Allah, the Most Gracious, the Most Merciful

I would like to express my deep gratitude to my supervisors Dr. Daniel Hampshire and Prof. Anne Goodeve for their continuous support throughout my PhD study and for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis.

I would also like to express my deep appreciation towards the members of the Haemostasis research group including Prof. Ian Peak, Dr. Martina Daly, Dr Vincenzo Leo, Dr Jacqui Stockley and Mr John Anson for their insightful comments and advices, but also for the hard question which encouraged me to widen my knowledge from various perspectives.

Special Thanks go to my colleagues, previous and current PhD student including Nasher Alyami, Ashley Cartwright, and Essa Sabi for their help and support during optimising and performing different experiments and techniques.

I am unable to describe in words how thankful I am to my Parents Hasan and Wedad, brothers and sisters and the rest of my family for their continuous support and encouragements, whom without them, I could have not completed this journey. They inspired me to find my dream and helped me to achieve it.

Last but not least, a heartfelt thanks goes to my wife Fatmah and my daughter Retal for their continuous support and help through difficult times during this research. Their patience and encouragements were undeniably a key in helping me finish my studies. Thank you, you have turned such a difficult times into enjoyable experience. I would like to thank my wife's family to staying beside us through these times and encouraging us to achieve the best we could from our studies.

Finally, Special thanks go to the Government of Saudi Arabia and Umm Al-Qura University in particular for their financial support.

Abstract

Background: von Willebrand factor (VWF) is a large multimeric plasma glycoprotein. It plays an important role in haemostasis by promoting platelet adhesion and aggregation at sites of blood vessel damage and by binding to and prolonging the half-life of coagulation factor VIII (FVIII). VWF is synthesised in endothelial cells where it undergoes post-translational modifications including glycosylation and multimerisation. VWF levels vary considerably in the normal population and several factors have been associated with this variation including ABO blood group and age. However, known factors only account for ~65% of variation in level. Previous studies suggest that other genetic loci, including *VWF*, *FUT3* and *CLEC4M* also influence VWF level. **Aims**: To identify genetic factors that influence VWF plasma level in the normal population and to investigate the mechanisms by which these genetic factors influence VWF level. **Methods**: DNA samples and phenotypic data (including VWF levels, age, gender and ABO blood group) were available for ~1100 healthy individuals (HC). Analysis was conducted on *VWF*, *FUT3*, *FUT5, FUT6* and *CLEC4M*. *In silico* analysis was used to identify variants predicted to affect expression or activity. Variants of interest were then genotyped in HC and their association with VWF level was investigated. *In vitro* studies were used to investigate the effect of these variants on VWF protein and mRNA expression, exon splice enhancer (ESE) motifs, splicing and mRNA half-life. **Results**: Investigation of 111 single nucleotide variants (SNV) in *VWF* identified 11 with significant association with VWF plasma level, of which c.2385T>C and c.2365A>G had been reported in previous studies. *In vitro* analysis showed that both were associated with increased protein, mRNA level and mRNA half-life. Investigation of variants in *FUT3*, *FUT5* and *FUT6* showed that only *FUT3* c.202T>C was significantly associated with VWF plasma level. Investigation of *CLEC4M* confirmed the association of a 23 amino acid variable number of tandem repeats (VNTR) with VWF plasma level. **Conclusion**: Variants within and outside *VWF* are found to be significantly associated with VWF plasma level through interfering with protein or mRNA levels. These factors were shown to have an additive effect and can be considered as risk factors for bleeding or cardiovascular disease.

Table of Contents

List of tables

List of figures

Publications and presentations arising from this thesis

Mufti AH, Ogiwara K, Swystun LL, Peake IR, Goodeve AC, Lillicrap D and Hampshire DJ on behalf of the EU-VWD and ZPMCB-VWD study groups. Single nucleotide variants c.2365A>G and c.2385T>C are associated with increased VWF plasma levels and VWF:FVIII binding. Poster presentation at the 2015 BSHT Annual Meeting, London, UK.

Mufti AH, Goodeve AC, Hampshire DJ on behalf of the EU-VWD & ZPMCB-VWD study groups (2015). Single nucleotide variants c.2365a>g and c.2385t>c are associated with increased VWF plasma levels through an effect on mRNA half-life. J Thromb Haemost **13 (Suppl. 2)**, abst. OR406. **The abstract received a Young Investigator Award from the ISTH**.

Mufti AH, Goodeve AC, Hampshire DJ on behalf of the EU-VWD and ZPMCB-VWD study groups. Single nucleotide variants rs1063857 and rs1063856 are associated with increased VWF plasma levels. Oral presentation at the 2014 BSHT Annual Meeting, Edinburgh, UK.

Mufti AH, Goodeve AC, Hampshire DJ on behalf of the EU-VWD and ZPMCB-VWD study groups. Single nucleotide variants rs1063857 and rs1063856 are associated with increased VWF plasma level. Poster presentation at the 8th Bari International Conference 2014, Bari, Italy.

Mufti AH, Lillicrap D, Peake IR, Goodeve AC, Hampshire DJ (2013). Investigation of the effect of *CLEC4M* on plasma von Willebrand factor level in the general population. J Thromb Haemost **11 (Suppl. 2)**, abst. PB 1.44-5

Mufti AH, Lillicrap D, Peake IR, Goodeve AC, Hampshire DJ on behalf of the EU-VWD & ZPMCB-VWD study groups. Investigation of the effect of CLEC4M on plasma von Willebrand factor level in the general population. Poster presentation at the Sheffield Medical school research day, Sheffield, UK.

List of abbreviations

+ve	positive
°C	Celsius
6-FAM	6-carboxyfluorescein
a.a	Amino acid
ABI	Applied Biosystems Inc
ABO	ABO blood group
ADAMTS13	A disintegrin and metalloproteinase with a thrombospondin type 1 motif 13
Ala	Alanine
Amp R	Ampicillin resistant
ANOVA	Analysis of variance
Approx	Approximately
APS	Ammonium persulfate
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ASSP	Alternative Splice Site Predictor
AT	annealing temperature
AVP	arginine vasopressin
B2M	Beta-2 microglobulin
BLAST	Basic local alignment search tool
bp	base pair
BRCA2	Breast and ovarian cancer gene 2
BSA	Bovine serum albumin
C	Cytosine
c.	Nucleotide alteration position using HGVS nomenclature
$CaCl_2$	Calcium chloride
cDNA	Complementary deoxyribonucleic acid
CHARGE	Cohorts for Heart and Aging Research in Genome Epidemiology
CK	Cysteine knot
CLEC4M	C-type lectin domain family 4 member M gene
cm	Centimetre
CNV	Copy number variation
CO_2	Carbon dioxide
CT	cycle threshold
Cys	Cysteine
dH_2O	Deionized water
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphates
DPBS	Dulbecco's PBS

dsDNA	Double stranded DNA
E.coli	Escherichia coli
EB	Elution buffer
Et Br	Ethidium bromide
EcoRI	E.coli endonuclease restriction enzyme
EDTA	Ethylenediamine tetra acetic acid
ELISA	Enzyme linked immunosorbent assay
ER	Endoplasmic reticulum
ESR1	Estrogen receptor gene
ER1	Estrogen receptor 1
ESS	Exonic splice silencer
EU	European Union
FBS	Foetal bovine serum
Fuc	Fucose
FUT1	Fucosyltransferase 1 gene
FUT2	Fucosyltransferase 2 gene
FUT3	Fucosyltransferase 3 gene
FUT5	Fucosyltransferase 5 gene
FUT6	Fucosyltransferase 6 gene
FVIII	Coagulation factor VIII
FVIII:C	Coagulation factor VIII activity
g	Gram
G	Guanine
g	G-force
GA	Golgi apparatus
GAIT	Genetic Analysis of Idiopathic Thrombophilia
Gal	Galactose
Glc	Glucose
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GpIbα	Glycoprotein Ibα
GPIIb-IIIa	Glycoprotein IIb-IIIa
GVGD	Align Grantham variation Grantham deviation
GWAS	Genome Wide Association Studies
h	Hour
H_2O_2	Hydrogen peroxide
HC	Healthy Control
HEK293T	Human embryonic kidney culture cell line
HEPES	Hydroxyethy-l-piperazine-1-ethanesulfonic acid sodium salt
Het	Heterozygous
His	Histidine
HMW	High molecular weight
Hom	Homozygous

HSF	Human splice finder
HWE	Hardy-Weinberg equilibrium
IC	Index case
Ile	Isoleucine
IU/dL	International units per decilitre
kb	Kilo base
kDa	Kilo Dalton
LB broth	Luria-Bertani broth
LD	Linkage disequilibrium
Lea	Lewis antigen a
Leb	Lewis antigen b
Leu	Leucine
Ltd.	Limited
Lys	Lysine
MCMDM-1VWD	Molecular and Clinical Markers for the Diagnosis and Management of Type 1 von Willebrand Disease
Met	Methionine
mg	Milligram
$MgCl_2$	Magnesium Chloride
Min	Minute
miRNA	MicroRNA
mL	Millilitre
mM	Milli molar
mRNA	Messenger RNA
Mvwf	Murine von Willebrand factor gene
n	Number
N/A	Not applicable
Nac	N-acetyl
NCBI	National Centre for Biotechnology Information
NetGene2	Splice site prediction tool
ng	Nanogram
NRG1	Neuregulin 1
NRG1	Neuregulin gene
OD	Optical density
OH	Hydroxide
OPD	o-Phenylenediamine
P	Partner
p.	Protein substitution position using HGVS nomenclature
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pH	Hydrogen ion
Phe	Phenylalanine
PIPES	Piperazine-N,N′-bis 2-ethanesulfonic acid
pmol	Pico mole
PolyPhen	Polymorphism phenotyping

pp	Pro-peptide
qRT-PCR	Quantitative real-time PCR
QTL	Quantitative trait loci
RNA	Ribonucleic acid
rpm	Rotations per minute
rs	Reference sequence
rVWF	Recombinant VWF
SCARA5	Scavenger receptor class A 5 gene
SDS	Sodium dodecyl sulphate
Sec	Second
Ser	Serine
SIFT	Sorting intolerant from tolerant
SNV	Single nucleotide variant
SP	Signal peptide
STAB2	Stabilin-2 gene
STX2	Syntaxin-2 gene
STXBP5	Syntaxin-binding protein 5 gene
SV40	Simian vacuolating virus 40
T	Thymine
Taq	Thermus aquaticus
TBE	Tris-borate-EDTA
TC2N	Tandem C2 domains nuclear protein gene
TEMED	Tetramethylethylenediamine
Thr	Threonine
Tm	Melting temperature
Tyr	Tyrosine
UCSC	University of California Santa Cruz
UK	United Kingdom
USA	United States of America
UTR	Untranslated region
UV	Ultra-violet
v	Volume
V2R	Arginine Vasopressin 2 receptor gene
Val	Valine
-ve	Negative
VNTR	Variable number of tandem repeats
VWD	von Willebrand disease
VWD1	von Willebrand disease type 1
VWF	von Willebrand factor
VWF	Von Willebrand factor gene
VWF:Ag	VWF antigen
VWF:RCo	Ristocetin co-factor activity
VWFP1	von Willebrand factor pseudogene
VWFpp	von Willebrand factor pro-peptide
VWFpp/VWF:Ag	Ratio of VWF pro-peptide to mature VWF

W	Weight
WPB	Weibel-Palade body
WT	Wild-type
α	Alpha
ß	Beta
ß-ME	ß- mercaptoethanol
µg	Microgram
µl	Microlitre
µM	Micro molar

1 Chapter 1: Introduction

1.1 Haemostasis

Haemostasis refers to the process of stopping the bleeding through blood clotting (conversion of blood from a liquid to a solid state) in order to repair an injured vessel and prevent loss of blood. The process involves several clotting factors that are involved in the clotting cascade, for example coagulation factor VIII (FVIII). Loss of function or reduction in level of any of these factors increases the risk of bleeding disorders such as haemophilia A, while increase of level increases the risk of different cardiovascular diseases such as venous thrombosis (Jenkins*, et al* 2012).

1.2 von Willebrand factor (VWF)

1.3 *VWF* and the VWF pseudogene

The VWF gene (*VWF)* maps to the short arm of chromosome 12 (12p13.3) (Ginsburg*, et al* 1985, Sadler*, et al* 1985, Verweij*, et al* 1985, Mancuso*, et al* 1989). It consists of 52 exons (51 coding; 2-17 encoding the pre-pro-peptide, 18-52 encoding mature VWF, Figure 1-1) (Mercier*, et al* 1991) transcribed to produce an 8.8kb mRNA product which is then translated to form a ~250kDa VWF protein (Mancuso*, et al* 1989, Sadler 1998). The *VWF* locus is known to be highly polymorphic with 8,463 single nucleotide variants (SNV) currently known to occur at this genetic locus (which is ~175kb in size) (Eikenboom*, et al* 1994; GeneCards V3, http://www.genecards.org, accessed December 2015).

There is also an ~21kb VWF pseudogene (*VWFP1*) located on chromosome 22q11.1 which shares ~97% homology with part of *VWF* (exons 23-34) (Mancuso*, et al* 1991). However, several sequence variants result in nonsense, missense, and splice site variation, which can also be used to distinguish between the *VWF* and *VWFP1* sequences (Mancuso*, et al* 1991).

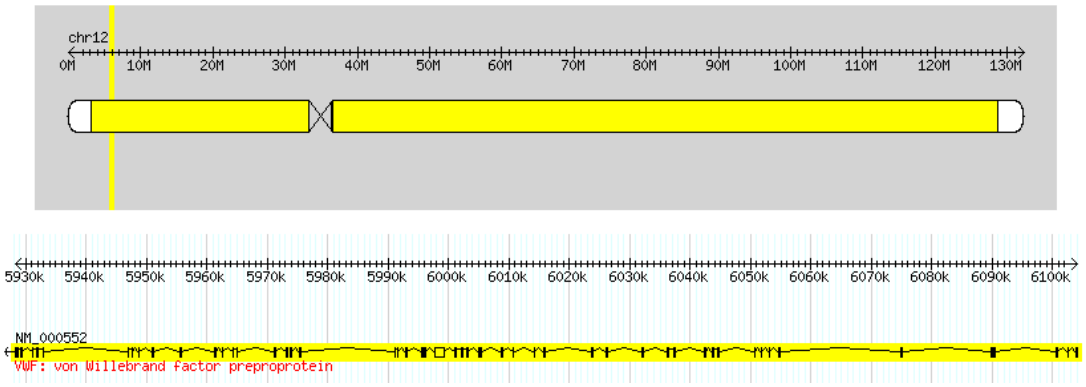


Figure 1-1: Location and structure of *VWF* gene (HapMap).

1.4 VWF protein

VWF is a large multimeric plasma glycoprotein which plays two central roles in the coagulation system (Jaffe*, et al* 1973, Sporn*, et al* 1985). First, VWF binds to circulating platelets facilitating their adhesion and aggregation to the vascular injury site as a result of high shear stress (Section 1.9) (Sadler 1998, Mohlke*, et al* 1999, Ruggeri 2001). Second, VWF binds to FVIII via the D' and D3 domains, protecting FVIII from proteolytic degradation which in turn extends the half-life of circulating FVIII (Montgomery*, et al* 1982, Kroner*, et al* 1991, Millar and Brown 2006, McKinnon*, et al* 2008). In addition, VWF carries FVIII, ensuring that FVIII is present at sites of vascular injury.

Historically, VWF structure consists of four types of conserved domains that are repeated throughout the protein (A, B, C and D) and a cysteine knot (CK) domain (Shelton-Inloes*, et al* 1986, Titani*, et al* 1986). Recent study has reviewed and updated the VWF sequence structure (figures 1-2 and 1-3) (Zhou*, et al* 2012). After the translation of the mRNA, it undergoes post-translational modifications including glycosylation and multimerisation (Titani*, et al* 1986, van Schooten*, et al* 2008). The final mature multimeric VWF is either secreted into the blood constitutively or stored in Weibel-Palade bodies in endothelial cells or in platelet α-granules (Wagner 1990).

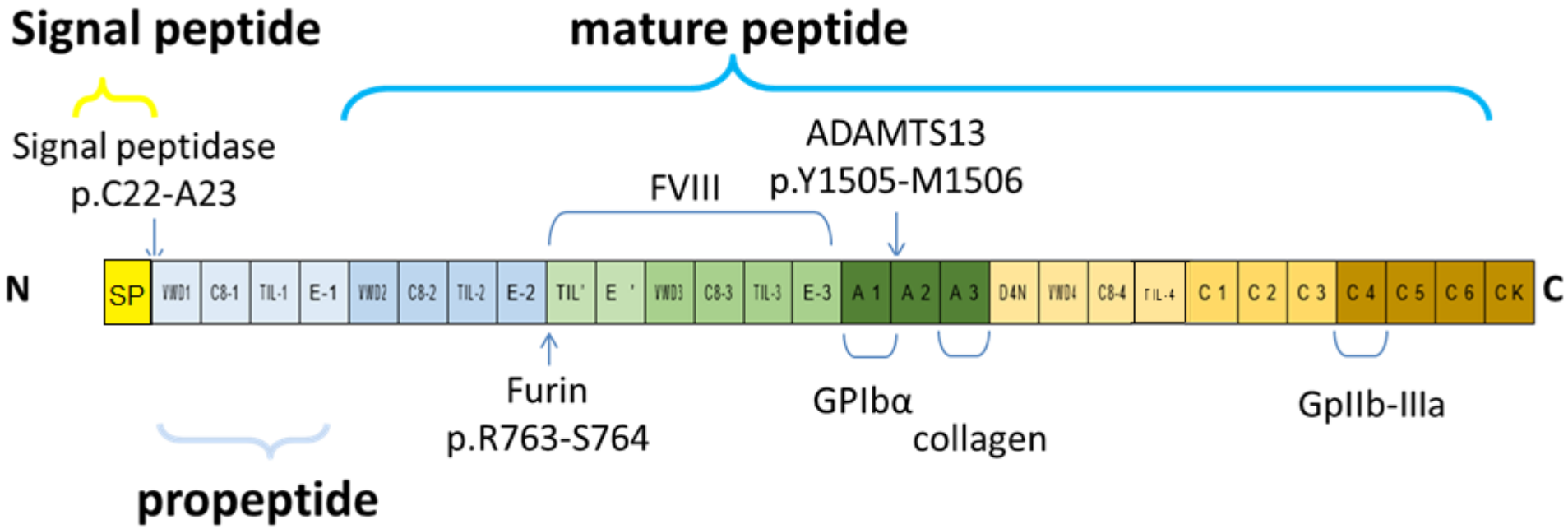


Figure 1-2: The translation product of *VWF.* Signal peptide highlighted in yellow; propeptide highlighted in light blue; mature peptide highlighted in darker blue and showing where different enzymes and compounds attach (Goodeve, *et al* 2007, Zhou, *et al* 2012).

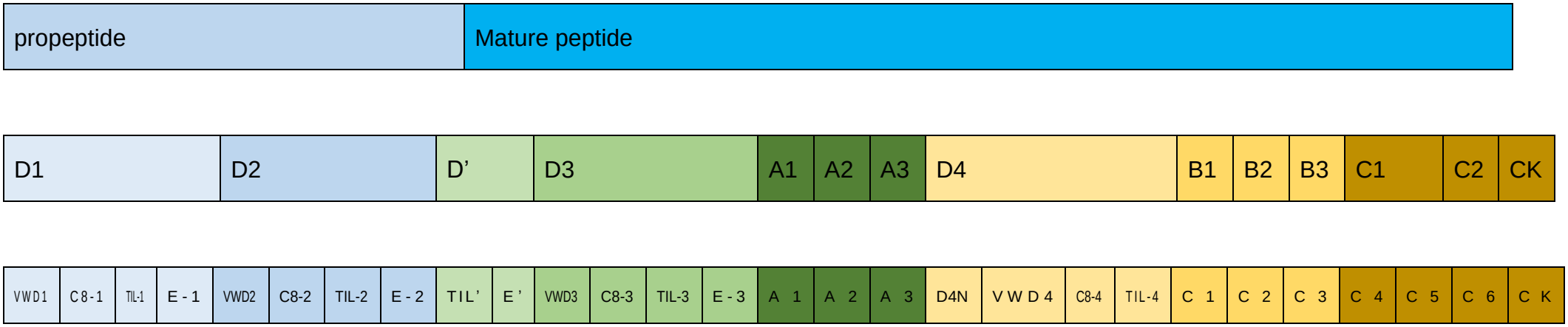


Figure 1-3: Diagram representing VWF protein structure. The diagram compares the old and new structuring of VWF protein showing each domain before and after recent structuring (Zhou, *et al* 2012).

1.5 Synthesis of VWF

The synthesis of VWF takes place in endothelial cells and megakaryocytes (Jaffe*, et al* 1973, Sporn*, et al* 1985, Wagner 1990). First, the monomeric VWF (pre-pro-VWF) is produced on the ribosome after the translation of VWF mRNA. The 2813 amino acid (a.a) pre pro VWF consists of a 22a.a signal peptide, a 741a.a propeptide and 2050a.a mature peptide. The pre-pro-VWF is translocated to the endoplasmic reticulum via the signal peptide (Sadler 1998, Sadler 2009).

In the endoplasmic reticulum (ER), pro-VWF (propeptide and mature peptide) is produced via cleavage of the signal peptide from pre-pro-VWF by signal peptidase (Figure 1-3). At this stage, the pro-VWF undergoes some post translation modifications including N-linked and O-linked glycosylation (for more detail, refer to section 1.6).

The monomeric VWF also undergoes dimerisation by disulphide bonding between the CK domains at the carboxyl terminals (between p.Cys2671 and p.Cys2813 residues) forming tail to tail binding between two VWF monomers (Figure 1-4). The 151a.a following the C6 domain are also reported to be of importance in the process of dimerization (Katsumi*, et al* 2000, Sadler 2009); a study expressing recombinant VWF lacking some of the a.a within this region showed ER retention of VWF due to its inability to form dimers (Yadegari*, et al* 2013). Following dimerisation, folding of VWF dimers is required before they are transferred to the Golgi apparatus for further post-translational modification. Incorrect folding could result in intracellular retention and proteasomal degradation (Yadegari*, et al* 2013). This suggests that disruption of VWF folding will result in intracellular retention and a reduction in the level of secreted VWF.

Once VWF dimers have been transported to the Golgi apparatus, further disulphide bonds bind dimers through the D3 domains to form a pro-VWF multimers (Wagner 1990, Sadler 2009, McKinnon*, et al* 2010). CxxC a.a motif sequences in the D1 and D2 domains are sites for protein disulphide isomerase activity which catalyses the formation of disulphide bonds between

the D3 and D' domains, which in turn enhances head-to-head multimer formation (Denis*, et al* 2008, McKinnon*, et al* 2010).

Finally, the propeptide bond (position p.Arg763-Ser764) is cleaved by Furin to produce the mature VWF which ranges in size from 2 to 200 monomers (the high molecular weight (HMW) form; Figure1-2) (Denis*, et al* 2008). The cleaved propeptide (VWFpp) remains bound to mature VWF via non-covalent bonds that are broken before VWF is secreted (Sadler 2009). VWFpp/VWF:Ag ratio can be used to assess the clearance rate of VWF from the circulation as an increased ratio suggests increased clearance rate (normal clearance when ratio is up to ~2.3 and enhanced clearance when ratio is >3) (Eikenboom*, et al* 2013).

The molecular weight of circulating VWF is also functionally important. Ultra-small VWF multimers are less able to recruit platelets to the site of vascular injury as they have fewer binding sites for platelets, while ultra large VWF multimers increase the risk of micro-vascular thrombosis as they have a very large number of binding sites, which might cause them to bind to platelets without prior vascular injury (Crawley*, et al* 2011). A disintegrin and metalloproteinase with a thrombospondin type 1 motif member 13 (ADAMTS13) performs the proteolytic cleavage of the ultra large VWF (Figure 1-2) (Furlan*, et al* 1996, Bowen and Collins 2004, Mannucci*, et al* 2004, Crawley*, et al* 2011).

Mutations within different VWF domains can in many cases explain the VWD phenotype. Mutations that may affect VWF level via a defect of the synthesis process are found mainly (but not only) in signalling peptide D or C domains (Gadisseur*, et al* 2009). Mutations in the signalling peptide can in many cases lead to intracellular retention as this domain is important in the translocation of VWF to the ER (Gadisseur*, et al* 2009). Mutations in the CK domain are likely to affect the dimerization process, thus leading to intracellular retention in the ER (Gadisseur*, et al* 2009). Mutation in D domains can lead to a defect in the multimerisation process therefore leading to intracellular retention as multimerisation is essential before VWF is secreted (Eikenboom*, et al* 1996).

On the other hand, mutations within these domains might not affect VWF level, but instead VWF function. There are many binding positions on VWF for different ligands (Figure 1-2), thus mutations affecting the binding may result in defects in VWF function instead of level.

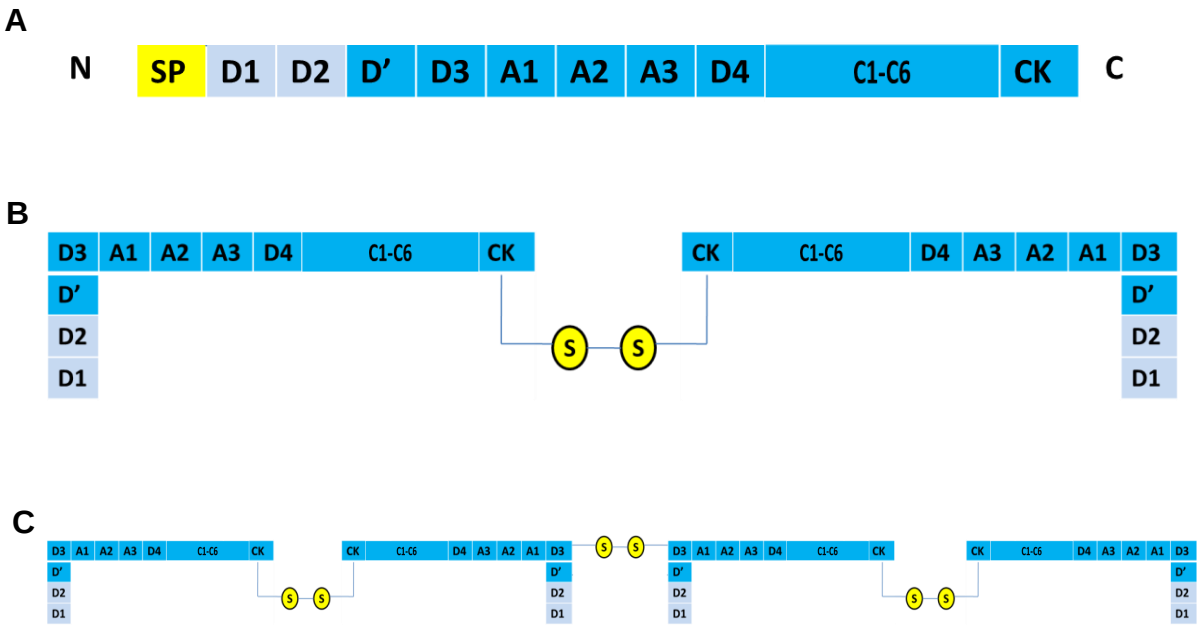


Figure 1-4: VWF; dimerization and multimerisation. The figure shows the process of VWF multimerisation. **A)** The translation product of *VWF* (the signal peptide in yellow, propeptide in light blue; mature peptide in blue) (Goodeve, *et al* 2007). **B)** Dimerization of VWF by disulphide bonding between the CK domains (Dong, *et al* 1994). **C)** The final multimeric VWF product formed by linking of VWF dimers via disulphide bonds between the D3 domains (Sadler 2009).

1.6 VWF glycosylation

Glycosylation refers to the addition of polysaccharide chains (glycans) to a protein. Glycosylation is a post-translational modification that occurs mostly in the endoplasmic reticulum and finishes in the Golgi apparatus (Matsui*, et al* 1993, Lenting*, et al* 2010, McKinnon*, et al* 2010). In total, the glycans added to VWF account for ~20% of the VWF molecular weight (McKinnon*, et al* 2010). It is reported that there are 10 O-linked glycans and 17 N-linked glycans (4 added to the propeptide and 13 to the mature sub-unit) added to each VWF monomer (Lynch*, et al* 1986, Titani*, et al* 1986, McKinnon*, et al* 2010).

Over 300 different structured glycans have been identified to date that can be added to VWF (Lenting*, et al* 2010, Lenting*, et al* 2015). Due to the small number of available glycosylation positions and the large number of available glycan structures, this results in large variation between each VWF monomer (Lenting*, et al* 2015). It is worth noting that some of the general features observed in these glycans, specifically that: 90% of glycans are sialylated including some of the O-linked glycans which have been found to be bi- or tri-sialylated glycans, five out of the 17 N-linked glycans have been found to be sulphated, ABH antigen can be found in ~13% of N-linked glycans and ~1% of O-linked glycans (Carew*, et al* 1990, Canis*, et al* 2010, Canis*, et al* 2012).

VWF glycosylation is found to have both structural and functional importance. It is essential for the formation of dimers (Wagner*, et al* 1986) and is also required for correct folding of VWF and promoting its structural stability (Lenting*, et al* 2010). In addition, it protects VWF from proteolytic and intracellular degradation, thus prolonging its half-life (Gill*, et al* 1987, Gallinaro*, et al* 2008). Moreover, glycosylation contributes to the clearance of VWF which in turn affects VWF half-life, as exemplified by the effect of the glycosylated ABH antigens on the clearance of VWF (section 1.12.1.2) (Millar and Brown 2006).

1.6.1 N-linked glycosylation

N-linked glycosylation occurs via the addition of glycans to asparagine residues in VWF. Of the 17 N-linked positions, 13% correspond to ABH glycosylation (Matsui*, et al* 1992, McKinnon*, et al* 2008). N-linked glycosylation affects the multimerisation of VWF (Wagner*, et al* 1986). Inhibition of N-linked glycosylation prevents VWF from transferring to the Golgi apparatus which in turn prevents the multimerisation process (Wagner*, et al* 1986).

A recent study investigated the importance of 17 N-linked glycans by mutating the Asn, which is required for glycosylation, to Gln at the binding site of each of the 17 N-linked glycans, preventing the attachment of one glycan at a time and studying the subsequent effect on plasma VWF (McKinnon*, et al* 2010). Mutation of glycosylation sites p.Asn99, p.Asn857 and p.Asn2790 was found to influence VWF plasma level. p.Asn99 was found to inhibit the production of VWF multimers completely, p.Asn857 was found to cause reduction of secreted VWF into the plasma and p.Asn2790 was found to be of importance for the dimerization process (McKinnon*, et al* 2010). Mutating the other N-linked glycosylation sites had a less significant impact on VWF production and secretion, and it was therefore suggested that they have other functional importance yet to be known (McKinnon*, et al* 2010).

1.6.2 O-linked glycosylation

Unlike N-linked glycosylation, little is known about O-linked glycosylation. O-linked glycosylation occurs via the addition of glycans to serine or threonine residues in VWF. There are 10 O-linked positions for glycosylation on VWF, of which sialyalated T-antigen was found to account for 70-90% while ABH antigen was found to account for around 1% (van Schooten*, et al* 2007, Canis*, et al* 2010, McGrath*, et al* 2010). It has been reported that O-linked glycans influence VWF half-life by an unknown process, although it has been suggested that they are associated with ADAMTS13 cleavage susceptibility (van Schooten*, et al* 2007).

1.7 VWF storage and secretion

Synthesised VWF is stored in α-granules in platelets and Weibel-Palade bodies (WPB) in endothelial cells (Wagner*, et al* 1982, Sporn*, et al* 1985). While the formation and packaging of VWF in platelets is still not clear, VWF is essential for the formation of WPB (Haberichter*, et al* 2005). Mutations leading to mis-folding of VWF or loss of multimeric VWF would lead to alteration in the formation and structure of WPB (Haberichter*, et al* 2005). WPB are suggested to initiate from the trans-Golgi network (Valentijn*, et al* 2011). VWF is compressed 100X in a tubule structure and stored in WPB before these structures are detached from the Golgi apparatus (Ferraro*, et al* 2014). This folding of VWF results in cigar-shaped WPB (Valentijn*, et al* 2008). The acidic environment in WPBs supports the self-assembly of VWF between the D'-D3 domains (Section 1.12.2.5) (Valentijn*, et al* 2008).

After the storage of VWF, platelets are known to release VWF upon activation (Nightingale and Cutler 2013). However, the process of release of VWF from WPB is not yet clear. Several suggestions have been proposed. The first and most widely acceptable is that WPB move randomly inside the cell and when they come close to the cell membrane, fusion occurs, resulting in the release of VWF into the circulation (Rondaij*, et al* 2006). However, the release of long strings of multimeric VWF was only observed upon stimulation *in vitro,* suggesting that other mechanisms are also involved (Nightingale and Cutler 2013, Lenting*, et al* 2015). Fusion of several WPB has also been observed resulting in the secretion of larger amounts of VWF after stimulation (Nightingale and Cutler 2013). Several proteins have also been associated with VWF secretion, including Rab15, Rab37, Ra1A, Munc18c, syntaxin-4, and other G proteins (Lenting*, et al* 2015).

1.8 VWF clearance

The half-life of the final multimeric plasma VWF is 12-20 hours (van Genderen*, et al* 1997). There are several factors influencing the half-life especially the size of the plasma VWF i.e. HMW-VWF has longer half-life as it takes longer to be cleared from the plasma (Goodeve 2010, Lenting*, et al* 2015). In addition, ABO blood groups was also reported to influence VWF half-life (Gallinaro*, et al* 2008). According to Denis *et al.* (2008), VWF is cleared by the liver and spleen; however, the clearance process is still not fully understood. It is also suggested that macrophages contribute to the clearance process (van Schooten*, et al* 2008, Rastegarlari*, et al* 2012, Lenting*, et al* 2015). In addition, it has been proposed that a glycoprotein receptor found on hepatocytes influences the clearance by initiating lysosomal degradation via interaction with the oligosaccharides found in VWF (Millar and Brown 2006).

Several studies have suggested factors that might contribute to VWF clearance, such as ABO blood group (section 1.12.1.2) and SNV within *VWF* (section 1.12.1.1).

Several receptors have been identified to be involved in clearing VWF from the circulation. The first identified receptor is called asialoglycoprotein receptor (Ashwell receptor) (Grewal*, et al* 2008). This receptor clears asialoglycoproteins from the circulation, but due to the fact that VWF is sialylated, this receptor is expected to only clear VWF in the absence/inactivity of sialyltransferase enzymes such as ST3Gal-IV (Ellies*, et al* 2002, Grewal*, et al* 2008).

Lipoprotein receptor LRP1, which is also expressed in macrophages, has also been associated with the clearance of VWF (Rastegarlari*, et al* 2012), having previously been known to be involved in the clearance of FVIII (Morange*, et al* 2005). This receptor better recognises unfolded VWF (through the A1 domain), therefore, it is of functional importance under shear stress (Wohner*, et al* 2015). In addition, a previous study showed that mice deficient in LRP1 had an increase in their VWF half-life (Rastegarlari*, et al* 2012). Furthermore, mutations in the *LRP1* have been associated with increased VWF plasma

levels in Brazilian normal individuals (Mello*, et al* 2008). A more recent study showed that mutation in LRP1 is associated with reduced VWF clearance and thus increased level (Wohner*, et al* 2015).

Siglec-5 is another receptor suggested to be involved in VWF clearance (Pegon*, et al* 2012). This receptor is expressed in macrophages supporting previous reports that macrophages play an important role in VWF clearance (van Schooten*, et al* 2008). However, the mechanism by which this receptor clears VWF is still not clear.

Recent genome wide association studies showed that *CLEC4M* is significantly associated with VWF plasma level (Smith*, et al* 2010). Further investigation associated L-SIGN, which is the product of *CLEC4M*, with clearance of VWF from the circulation (Rydz*, et al* 2013). This receptor is reported to be expressed in sinusoidal endothelial cells in the liver, which further supports its association with VWF clearance.

1.9 VWF function

As previously mentioned, VWF has two main functions: 1) binding to circulating FVIII and protecting it from proteolytic degradation, therefore prolonging its life & delivering to site of vascular injury, and 2) binding and recruiting platelets to the site of vascular injury (Baruch*, et al* 1991, Ruggeri 2001).

VWF binds to FVIII via domains D'-D3, therefore mutations within this region can result in the inability of VWF to bind to FVIII which in turn results in reduction of FVIII level (figure 1-2 and figure 1-5) (Nishino*, et al* 1989, Sadler*, et al* 2006). This leads to type 2N VWD (section 1.11.1) which exhibits a mild haemophilia A-like phenotype.

VWF recruits platelets to the site of injury, helping to form a blood clot which prevents further bleeding. In a normal scenario, VWF multimers circulate in a folded state, reducing the number of binding sites for platelets in order to prevent spontaneous binding of platelets to VWF which might cause a blood clot within the small arteries (arterioles) or veins (venules) (Sadler 1998,

Ruggeri 2001). However, after vascular injury, VWF binds to collagen (normally present in the sub-endothelial matrix) through the A3 and A1 domains, which causes VWF to linearize with the help of shear stress from the flow of blood (Reininger 2008). This in turn exposes more platelet binding sites in the VWF multimers which allows more platelets to bind to VWF through the platelet receptor glycoprotein Ibα (GpIbα) receptor (Savage*, et al* 1996, Reininger 2008). Bound platelets become activated, change in shape and begin to express the glycoprotein IIb-IIIa (GPIIb-IIIa) receptor which rapidly binds to VWF through domain C4 (Keeney and Cumming 2001). Further circulating VWF binds to activated platelets, becomes linearized by shear stress leading to recruitment of further platelets and eventually to formation of a thrombus (figure 1-5) (Kenny*, et al* 2002).

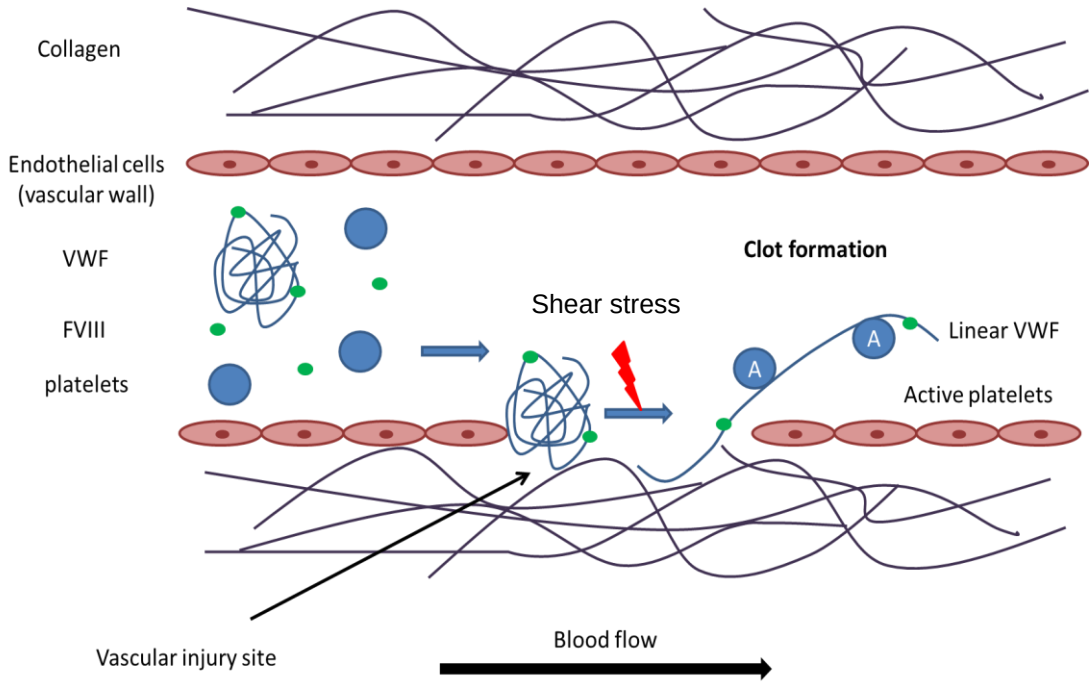


Figure 1-5: VWF interaction with platelets and FVIII at the site of vascular injury. Following vascular injury, VWF binds to collagen (via the A1 and A3 domains) and undergoes conformational changes due to high shear stress which changes its form from folded to linear and leads to the binding of platelet receptor GpIbα via the A1 domain (Fujimura*, et al* 1986, Hoylaerts*, et al* 1997, Mazzucato*, et al* 1999).

1.10 Types of VWF

There are two forms of VWF, plasma VWF and platelet VWF. Both plasma and platelet VWF are encoded by *VWF,* however, plasma VWF is produced in endothelial cells while platelet VWF is produced by megakaryocytes. Both types of VWF are found to have significant differences in their structure and glycosylation pattern (Denis*, et al* 2008, McGrath*, et al* 2010). Plasma VWF is found to go through more extensive proteolysis compared to platelet VWF (McGrath*, et al* 2010). Storage is also different, with plasma VWF stored in Weibel-Palade bodies and platelet VWF stored in α-granules (Ruggeri and Zimmerman 1980, Wagner 1990). Platelet VWF is also found not to bind and carry FVIII in the circulation as it is suggested to be permanently stored and not released from the platelets in mice (Yarovoi*, et al* 2003). The majority of investigations, including this study focus on plasma VWF.

1.11 VWF level

Normal plasma levels of VWF range between 50-200 IU/dL in 95% of the general population and therefore shows a wide range compared to most other coagulation factors such as FII, FV, FIX, FX, FXI and FXII where the normal level ranges between 70-130 IU/dL (Gill*, et al* 1987, Mannucci*, et al* 1989, Iverson*, et al* 2007). High VWF levels (>200 IU/dL) are associated with increased thrombotic risk, myocardial infarction and stroke (Galbusera*, et al* 1997, Lemmerhirt*, et al* 2006, Morga-Ramirez*, et al* 2010, van Schie*, et al* 2011a, van Schie*, et al* 2011d) and low VWF levels (<50 IU/dL) are associated with the bleeding disorder von Willebrand disease (VWD) (Lemmerhirt*, et al* 2006, Mazzeffi and Stone 2011).

1.11.1 von Willebrand disease (VWD)

VWD was first reported in 1926 by Finnish physician Dr Erik von Willebrand. VWD is the most common inherited bleeding disorder and is caused by either a quantitative or qualitative deficiency in VWF mainly as a result of *VWF* mutation (Sadler 2004, Sadler*, et al* 2006, Nichols*, et al* 2008, Lenting*, et al* 2012). It results in various symptoms including bleeding after surgery or tooth

extraction, subcutaneous bleeding and increased risk of epistaxis (Rodeghiero*, et al* 1987, Sadler*, et al* 2006, Tosetto*, et al* 2006).

There are three main types of VWD; type 1, type 2 and type 3. Type 1 (VWD1) is the most common and accounts for around 70% of all VWD cases. It is caused by a reduction in plasma VWF level and is characterised by a mild/moderate bleeding phenotype (Eikenboom*, et al* 2006). VWD1 is generally inherited in an autosomal dominant manner. VWD1 patients with the same mutations have been shown to have variable VWF plasma levels and VWD1 can also demonstrate incomplete penetrance, i.e. an individual with a known VWD1 mutation might not exhibit symptoms (Cumming*, et al* 2006). Different types of mutations have been reported to result in VWD1 including missense, nonsense and splice site mutations.

Type 3 is the rarest and most severe type of VWD, occurring in 5% of all VWD patients, and is characterised by almost a complete lack of plasma VWF (Sadler*, et al* 2006). It has autosomal recessive inheritance. The most common causative mutations for this type are nonsense mutations which account for ~30%, although other mutations including deletions, insertions and splice site mutations can also be causative.

Type 2 VWD is caused by defects of VWF function and is divided into four subtypes; 2A, 2B, 2M and 2N. Type 2A occurs in a reduction in the HMW VWF (Enayat*, et al* 2001). It is mostly inherited in an autosomal dominant manner with mutations resulting in loss of HMW VWF either due to enhanced ADAMTS13 cleavage or a defect in the dimerization or multimerisation processes (Goodeve 2010). Type 2B is inherited in an autosomal dominant manner and occurs in an increased VWF binding affinity to platelet GpIbα which is due to missense mutations in the A1 domain (Sadler*, et al* 2006, Lenting*, et al* 2012). Type 2M is also inherited in an autosomal dominant manner and occurs from reduced binding of VWF to platelets (Lenting*, et al* 2012). Type 2N is inherited in an autosomal recessive manner and occurs in a decreased binding affinity of VWF toward FVIII (Sadler 1998, Sadler*, et al* 2006).

Given that mild bleeding symptoms are common even in the normal population, it is hard to distinguish between patients with mild VWD1 and normal healthy individuals. Due to these factors, the prevalence of VWD1 patients diagnosed with VWD is approximately 1/10,000, while the actual prevalence could be much higher (Lillicrap 2007). In addition, several VWD1 studies failed to identify mutations in a subset of their patients (Goodeve*, et al* 2007, James*, et al* 2007), suggesting that other genetic factors are expected to contribute to a VWD-like phenotype.

1.12 Factors influencing VWF plasma level

Due to the fact that several genetic and environmental factors are reported to have an effect on VWF level, there is no accurate distribution of these factors with VWF level (Figure 1-6).

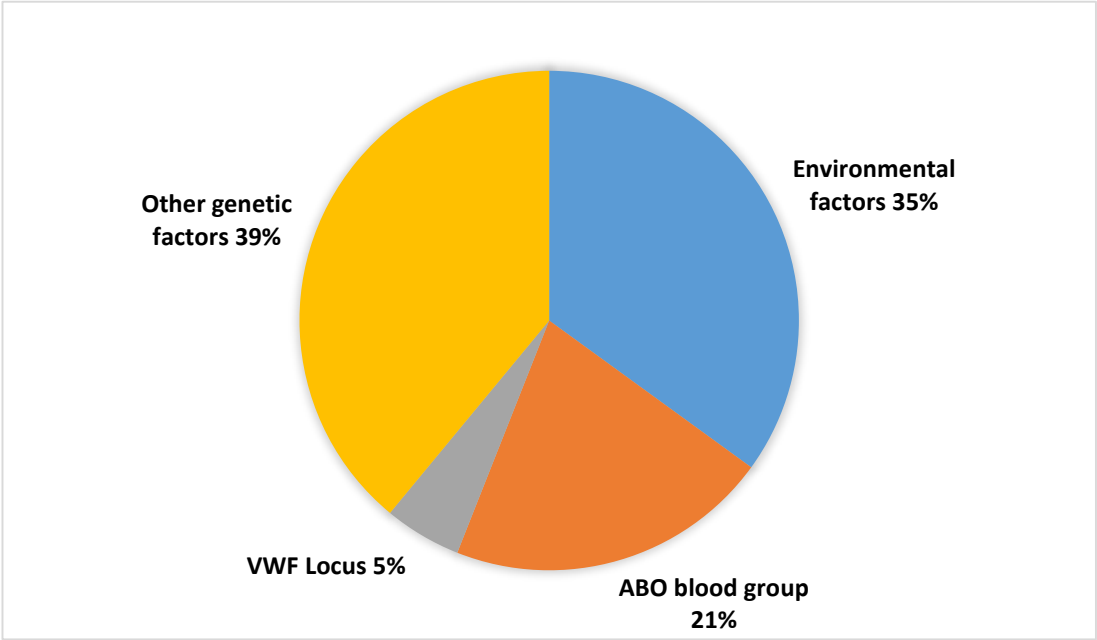


Figure 1-6: Pie chart showing factors influencing VWF plasma level with their extent of contribution in the general population (Desch 2012).

1.12.1 Genetic factors

1.12.1.1 VWF

Polymorphic variation within and adjacent to *VWF* has been shown to influence VWF level. *VWF* is a quantitative trait locus (QTL) for VWF levels and is responsible for ~20% variation in VWF plasma level after accounting for ABO blood group (De Visser*, et al* 2003). This includes the $(GT)_n$ repeat in the promoter region and the SNV c.2771G>A (p.Arg924Gln) (Hough*, et al* 2008, Hickson*, et al* 2010, Hickson*, et al* 2011). The $(GT)_n$ repeat has been shown to affect *VWF* promoter activity and contributes to variation in VWF levels by around 10%, with longer repeats found to have higher *VWF* promoter activity (Hough*, et al* 2008, Hickson*, et al* 2011), which contradicts the results of another smaller study (Daidone*, et al* 2009).

Studies have also identified four SNVs located in the promoter region of *VWF* associated with plasma VWF level; c.-3268G>C, c.-2709C>T, c.-2661A>G and c.-2527G>A (Zhang*, et al* 1994, Keightley*, et al* 1999, Harvey*, et al* 2000). The association of these SNV was more significant in individuals over 40 years old. Using some *in silico* analysis, these SNV were predicted to potentially interact with several transcription factors such as GATA-2, c-Ets and Sp1, which may play a role in regulating *VWF* expression (Keightley*, et al* 1999, Harvey*, et al* 2000). However, a more recent study which investigated a larger number of individuals showed no significant association between c.-2527G>A and VWF plasma level, which can also be applied to the other three SNVs because all four SNVs are in complete linkage disequilibrium (LD) (Hickson*, et al* 2011) suggesting that these variants might be in partially strong LD with another variant that is responsible for the observed association.

1.12.1.2 ABO blood group

Gill *et al.* (1987) was the first to report an association between ABO blood group and VWF plasma level. There are three main antigens in the ABO system; A, B and H. A antigen is found in people with blood group A, B antigen is found in people with blood group B, both A and B antigens are found in

people with blood group AB, whereas H antigen is found in people with blood group O.

The synthesis of different ABH antigens starts with a basic precursor (glycan chain) (Figure 1-7). First, the precursor is converted to H antigen by α-2-fucosyltransferase which is encoded by fucosyltrasferase 1 (*FUT1*) (O'Donnell*, et al* 2005). H antigen is then converted to A or B antigens via a glycosyltransferase which is encoded by the *ABO* gene. *ABO* encoded N-acetyl-D-galactosamine transferase adds N-acetyl-D-galactosamine to H antigen converting it to A antigen, while *ABO* encoded D-galactose transferase adds galactose to H antigen converting it to B antigen (Klarmann*, et al* 2010). *ABO* encodes three main allelic forms of which A and B encode different forms of glycosyltransferase while O encodes an inactive form of glycosyltransferase resulting from a single nucleotide deletion (Figure 1-7) (Dean 2005).

Individuals with different ABO blood groups are found to have different VWF levels (Gill*, et al* 1987, O'Donnell*, et al* 2005). Individuals with more complex ABH antigens (i.e. AB blood group individuals) have a longer VWF half-life compared to individuals with less complex ABH antigens (i.e. O blood group individuals) who have on average ~25% lower VWF level compared to non O blood group individuals (Figure 1-8) (Gill*, et al* 1987, Gallinaro*, et al* 2008, Castaman*, et al* 2010, McGrath*, et al* 2010). More complex antigens cause VWF to be less susceptible to clearance from the circulation (Gallinaro*, et al* 2008, Castaman*, et al* 2010).

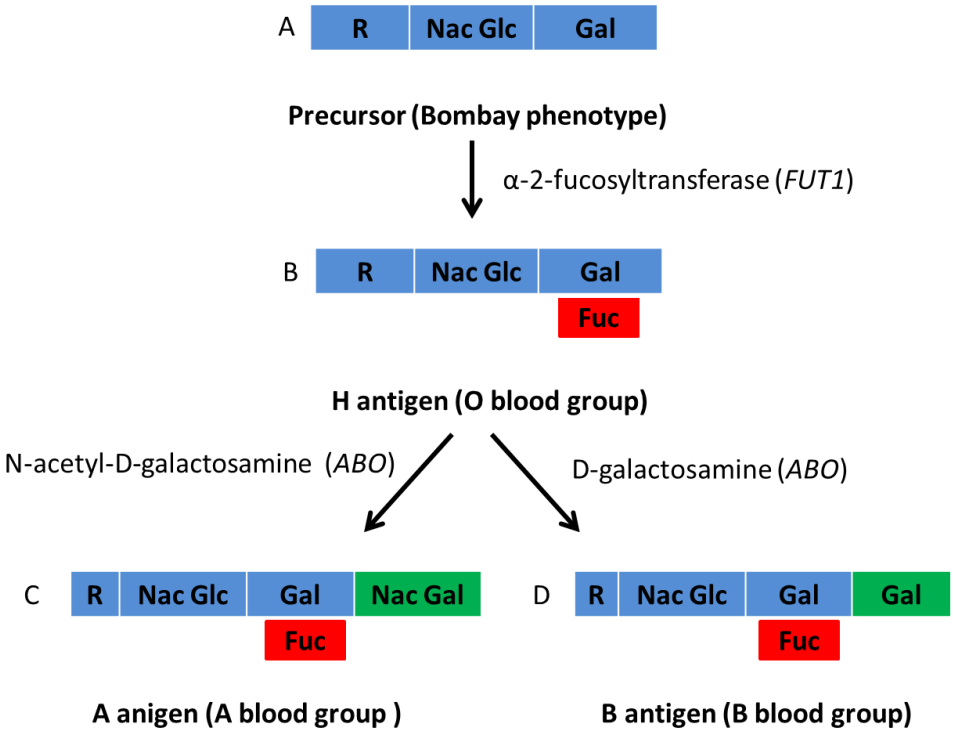


Figure 1-7: Synthesis of ABH antigens. The figure shows the synthesis of different ABH antigens (starting with the basic precursor glycan) by the required enzymes. It shows the precursor glycan **(A)**, H antigen **(B)**, A antigen **(C)** and B antigen **(D)** (Klarmann*, et al* 2010). (Glc, glucose; Gal, galactose; Nac, N-acetyl; Fuc, Fucose; R, the rest of the glycan chain).

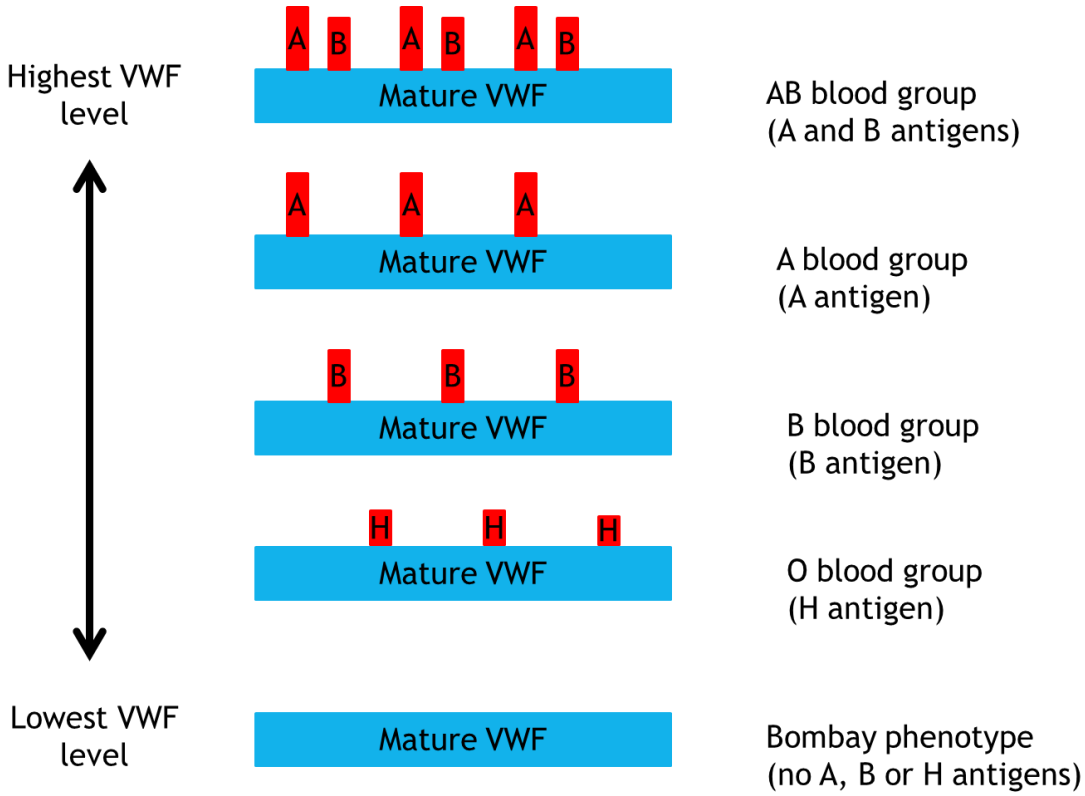


Figure 1-8: The effect of different ABH antigens on VWF plasma level. Individuals with AB blood group are found to have highest VWF plasma levels (average 123.3 IU/dL), followed by individuals with A blood group (116.9 IU/dL), then B blood group (105.9 IU/dL) and O blood group (74.8IU /dL). Individuals with Bombay blood group are found to have the lowest VWF plasma levels (~10% lower than O-blood group individuals) (Gill, *et al* 1987, O'Donnell, *et al* 2005).

1.12.1.3 Bombay and para-Bombay blood groups

Individuals with a Bombay blood group lack all ABH antigens in their plasma and exocrine secretions due to their inability to express FUT1 and FUT2 as a result of null mutations in both encoding genes (O'Donnell*, et al* 2005). On the other hand, Individuals with a para-Bombay blood group have H antigen in their exocrine secretion due to their ability to express FUT2, while they lack the H antigen in their plasma due to their inability to express FUT1 (O'Donnell*, et al* 2005). Recent studies reported that individuals with Bombay or para-Bombay blood groups have 10-15% lower VWF plasma levels compared to individuals with O blood group as a result of increased proteolysis susceptibility by ADAMTS13 (O'Donnell*, et al* 2005, Klarmann*, et al* 2010).

1.12.1.4 Arginine vasopressin 2 receptor (V2R)

Recent research was conducted on normal individuals to study the association between V2R and VWF plasma level because V2R is responsible for maintaining blood pressure and volume, and is expressed by endothelial cells. The study found that V2R is associated with ~10% VWF level (Nossent*, et al* 2010). V2R regulates the secretion of VWF into plasma by Weibel-Palade bodies. Mutations (i.e. p.Gly12Glu, p.Ala245Cys and p.Ser331Ser) in *AVPR2,* the gene encoding V2R mapped to the X chromosome, leads to a high binding affinity between V2R and arginine vasopressin (AVP), which increases signals transduction, resulting in increased excretion of VWF from Weibel-Palade bodies (Kaufmann*, et al* 2000, Nossent*, et al* 2010). However, the association of V2R was not observed in association studies such as the CHARGE and GAIT studies (Souto*, et al* 2003, Smith*, et al* 2010), which may be due to V2R having a weak association with VWF level.

1.12.1.5 Genome-wide association studies

Several genome-wide association studies (GWAS) have been conducted in order to find novel genetic loci associated with VWF plasma levels.

The Cohort for Heart and Aging Research in Genome Epidemiology (CHARGE) study, examined 23,608 Caucasian individuals from 5 centres (Smith*, et al* 2010). The study scanned the participants for SNVs associated

with VWF level, FVIII and FVII activity. Eight different genetic loci were associated with VWF levels, including *VWF* and *ABO* (Smith*, et al* 2010). The other genetic loci were *STXBP5*, *SCARA5*, *STAB2*, *STX2*, *TC2N* and *CLEC4M*, none of which was previously reported to be associated with VWF level (Smith*, et al* 2010). These genes are suggested to be of functional importance toward VWF level as they were suggested to be associated with transport and cell membrane receptors, which may influence VWF synthesis, storage or clearance (Smith*, et al* 2010).

The Genetic Analysis of Idiopathic Thrombophilia (GAIT) study identified several genetic loci associated with VWF plasma level in 21 Spanish families: 1p36.13, 2q23.2, 5q31.1, 6p22.3, 9q34.2 and 22q11.1 (Souto*, et al* 2003). None of the identified genetic loci were reported by the CHARGE study except for *ABO* (9q34.2). However, this study failed to identify specific genes that might be responsible for this association due to the small sample number used (n= 398) and the lack of extended family pedigrees.

Recent association studies further confirmed the association of *ABO* and *VWF* with VWF plasma level (Campos*, et al* 2011, Zabaneh*, et al* 2011). Campos and colleagues (2011) studied the association of different SNV and haplotypes within *VWF* with VWF plasma level and identified 18 SNV with significant association, all of which are present within a 50kb region covering exons 13-24 which encodes the D2, D' and D3 domains. One of these SNVs (rs1063857 in exon 18) was identified previously by the CHARGE study (Smith*, et al* 2010). In addition, they identified 9 different haplotypes from the previous SNVs with significant association with VWF plasma level. On the other hand, Zabaneh and colleagues (2011) studied the association of different genetic loci with VWF plasma level. This study identified 2 new genetic loci with a weak association to VWF plasma level; estrogen receptor 1 (*ESR1*) and neuregulin 1 (*NRG1*). Not much is known about *ESR1* in relation to VWF level, however, *NRG1* is reported to be expressed in plasma by endothelial cells and has been associated with chronic heart failure (Ky*, et al* 2011).

1.12.1.6 Animal studies

Several studies have been conducted on inbred mice of two different strains, RIIIS/J, which is a strain that exhibits a VWD1 phenotype with prolonged bleeding and low VWF and FVIII level, and PWK/Ph, which is a strain with a normal VWF level (Nichols*, et al* 1994). These studies have found seven genetic loci associated with VWF plasma level, *Mvwf1-7* (Mohlke*, et al* 1996, Lemmerhirt*, et al* 2006, Lemmerhirt*, et al* 2007, Shavit*, et al* 2009), all of which occur in locations distinct from *Vwf* except for *Mvwf2* and *Mvwf5* (Shavit*, et al* 2009). Mohlke and colleagues (1996) identified a locus that is associated with a 20-fold decrease in plasma VWF level, which occurs on chromosome 11, and named it *Mvwf1* (Mohlke*, et al* 1996). It is believed that a gain of function mutation at this locus is responsible for this effect on VWF plasma level (Mohlke*, et al* 1999). Further study on this locus led to the identification of the gene *Galgt2* (Mohlke*, et al* 1999). It is suggested that the *Mvwf1* human Ortholog (*B4GALNT2*) may have similar associations with VWF plasma level (Mohlke*, et al* 1996, Mohlke*, et al* 1999, Dall'Olio*, et al* 2014).

Mvwf2 mapped to mouse *Vwf* and a p.Arg2657Gln (c.7970G>A) mutation at this locus was found to affect the secretory pathway of VWF by influencing the synthesis and secretion of VWF (Lemmerhirt*, et al* 2006). *Mvwf2* was found to be responsible for around 29% of the variation in mouse VWF level (Lemmerhirt*, et al* 2007).

A variant in *Mvwf5* was found to affect the binding of regulatory components in Vwf itself such as transcription regulation or RNA stability, which in turn affects the expression of Vwf (Shavit*, et al* 2009). Although the mechanisms associated with *Mvwf3*, *Mvwf4*, *Mvwf6* and *Mvwf7* are yet to be known, a potential human orthologue for *Mvwf6* was reported in the GAIT study to be present within 12q12 and 21q22.3, which might suggest that these loci are also important in influencing *VWF* plasma level (Souto, et al 2003, Shavit, et al 2009).

1.12.2 Environmental factors

1.12.2.1 Ethnic background

Research conducted by Kadir and colleagues (1999) found that black women tend to have significantly higher VWF levels in comparison to other ethnic groups. This result confirmed several previous studies which found that black individuals had higher VWF levels than white individuals (Meade*, et al* 1977, Meade*, et al* 1978, Conlan*, et al* 1993). This difference in VWF level was thought to be due to differences in living habits such as smoking habits, and genetic background (i.e. genomic variation) in individuals from different ethnic backgrounds (Meade*, et al* 1978).

1.12.2.2 Gender

The effect of gender on VWF plasma level has been a controversial topic. Some studies suggest that males and females do not differ in VWF plasma level, while other studies found that males have higher VWF levels than females, and others believe that females have higher VWF plasma levels compared to males (Conlan*, et al* 1993). However, another study aimed to reassess the effect of gender on VWF plasma level and suggested that there are no significant differences in levels between males and females (Favaloro*, et al* 2005) which was supported by a recent study (Zhou*, et al* 2014).

1.12.2.3 Age

Several studies have shown an association of age with plasma VWF level. Individuals exceeding 40 years old are found to have increased VWF plasma levels (Chakrabarti*, et al* 1975, Conlan*, et al* 1993, Kadir*, et al* 1999, Konkle 2014) and 4.8% of variation in VWF plasma level is reported to be due to the effect of age (Campos*, et al* 2011). It has also been reported that adults have an average 10 IU/dL VWF level increase per decade (Laffan*, et al* 2004). Another study conducted on young children found that after 1-3 months, VWF plasma level decreased until the age of 13 months to 4 years, when the VWF plasma level then starts to increase again (Klarmann*, et al* 2010).

1.12.2.4 Estrogen level

In vitro, estrogen is found to increase secretion of VWF by endothelial cells. (Harrison and McKee 1984). However, this observation has not been tested *in vivo*, which might make these findings questionable.

1.12.2.5 Golgi apparatus pH level

Wagner and colleagues (1986) suggested that an acidic environment is essential for the formation of multimeric VWF inside the Golgi apparatus. A recent review by Sadler (2009) reported that a basic environment inside the Golgi apparatus inhibits the formation of the disulphide bonds required for VWF multimerisation, thereby lowering the production of HMW-VWF, which is in line with previous findings (Wagner*, et al* 1986).

1.13 Aims and hypothesis

The main goal of this project is to identify novel genetic factors that may contribute to the wide variation in VWF plasma levels (50-200 IU/dL) reported in the general population; these factors may also explain the difference in VWD/thrombosis severity in patients with the same causative mutations. Therefore, the aims of this project are:

1. To identify novel genetic loci associated with VWF plasma level
2. To investigate the mechanism(s) by which these genetic loci influence VWF plasma level

Given that a large number of genes have been suggested to be associated with VWF plasma level, this study initially focused on those genes strongly associated with VWF levels for further laboratory investigation: *VWF*, *FUT3* and *CLEC4M*. It is hypothesised that variants within and outside VWF are associated with VWF plasma level in the normal population.

2 Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Study population

The study population consisted of 1166 healthy control individuals (HC) who were recruited to the European Union study, Molecular and Clinical Markers for the Diagnosis and Management of Type 1 von Willebrand Disease (MCMDM-1VWD), by 12 European partners (P): P1) Sheffield, UK; P2) Vicenza, Italy; P3) Milan, Italy; P4) La Coruña, Spain; P5) Paris and Lille, France; P6) Leiden, the Netherlands; P7) Hamburg, Germany; P8) Aarhus, Denmark; P9) Prague, Czech Republic; P10) Malmo and Lund, Sweden; P11) Leicester, UK and P12) Birmingham, UK. Ethical review and informed consent were obtained prior to study participation.

Phenotypic data including age, gender, ABO blood group, von Willebrand factor (VWF) plasma antigen levels (VWF:Ag), ristocetin cofactor activity (VWF:RCo), VWF propeptide (VWFpp) plasma level and factor VIII activity (FVIII:C), were available for the majority of HC (Goodeve*, et al* 2007, Eikenboom*, et al* 2013). Caucasian was the ethnicity of the majority of the recruited individuals. Genomic DNA from HC was previously extracted from citrated whole blood using standard procedures and stored at -20°C.

2.1.2 Chemicals, reagents and commercial kits

Chemicals and reagents used in this project were obtained from various sources. OneTaq DNA polymerase and Q5 high fidelity DNA polymerase were obtained from New England Biolabs Ltd. (Hitchin, UK). KAPA HiFi polymerase was obtained from KAPA Biosystems Ltd. (London, UK). Lipofectamine LTX (transfection reagent) was obtained from Invitrogen by Life Technologies (Ltd. Paisley, UK). 5× passive lysis buffer (5× PLB) and Renilla Luciferase Assay System were obtained from Promega Ltd. (Southampton, UK). All restriction enzymes and related buffers were obtained from New England Biolabs Ltd. Antibodies used for ELISA were obtained from Enzyme Research (Laboratories Ltd. Swansea, UK). ELISA standard VWF:Ag calibrator was obtained from Quadratech (Diagnostics Ltd. Epsom, UK). Hyperladder molecular weight markers and agarose powder, which were used for

performing gel electrophoresis were obtained from Bioline Limited (London, UK).

QIAquick Gel Extraction Kit, which was used to extract DNA from agarose gel, EndoFree plasmid maxi and QIAprep Spin Miniprep kits (which were used to extract DNA from competent cells) were all obtained from QIAGEN Ltd. (Crawley, UK). Quikchange Lightning Site Directed Mutagenesis kit and XL10-gold *E-coli* were obtained from Agilent Technologies Ltd. (UK, Stockport). Luria Bertoni (LB) and NZY broth and agar were obtained from Merck (Darmstadt, Germany).

Dulbecco's Modified Eagles cell culture Media (DMEM), foetal bovine serum (FBS) and 10x phosphate buffered saline (PBS) were obtained from Life Technologies.

2.1.3 Primers

The oligonucleotide primers used in this study (Tables 2-1, 2-2, 2-3 and 2-4) were ordered lyophilised from Eurofins MWG Operon (Ebersburg, Germany) and on arrival were diluted to 100pmol/µl according to the manufacturer's guidelines. Refer to section 2.2.1.5 for primer design and validation.

Table 2-1: PCR primers used for amplification and sequencing in this study.

Primer	Sequence (5' to 3')
FUT3F	M13-TTACCTGCAAAACCCAGTCC[1]
FUT3R	M13-TCAGTGTGGCAAGGTCTCTG[1]
FUT5F	M13-GCACTGGCTCTTTCCTCATC[1]
FUT5R	M13-AGTAGGTGCAGCCTCCAGAA[1]
FUT6F	M13-GACCCTTGCCCATGTAGAAA[1]
FUT6R	M13-TCAGCCAGAACCATCACTCA[1]
FUTseqF1	AATCTCACCATGTCCTACCG
FUTseqR1	CGTACACGTCCACCTTGAGA
FUTseqF2	TGGCTGTGTGTTTCTTCTCC
FUTseqR2	TTGTCCAGCTCCTGCAGGTA
CLEC4MrepF	FAM-GAGACCTTGGCTCTCACAAA[2]
CLEC4MrepR	GTGTTCTCATTTCACAGGGG
CLEC4M_SNV_F	M13-GGTGAGTTGCCAGACCAGTC[1]
CLEC4M_SNV_R	M13-GTGTACCTGCTCCTCAGCAG[1]
CLEC4M_7_F	M13-GCAGGAACTGAAGGCAACAT
CLEC4M_7_R	M13-GGAAACTTGGAGGCATCAGA
FUT3_expression_F	ACATGGCCTTTCCACATCC
FUT3_expression_R	AGGCGTGACTTAGGGTTGG
CLEC_4M_expression_F	AAGTCAGGAACAATCCG
CLEC_4M_expression_R1	ACTGGTCTGGCAACTCA
CLEC_4M_expression_R2	TTGGAAGAATGTCCAGTC
pcDNA-Dupseq	CACTCAACCCTATCTCGGTCTA
FUT3_qPCR_F	AGCCAAGCCACAATGGCCAT
FUT3_qPCR_R	GGGTTGGACATGATATCCCAG

[1]M13F (5'-TGTAAAACGACGGCCAGT-3') or M13R (5'-CAGGAAACAGCTATGACC-3') were attached in order to facilitate sequencing; [2]a 6-carboxyfluorescein (6-FAM) fluorescent tag was attached to facilitate accurate sizing of amplified products.

Table 2-2: Mutagenic primers.

Variant	Sequence (5' to 3')
a2365g	ctgaagggctcgagtgtgccaaaacgtgcc
a2365g_antisense	ggcacgttttggcacactcgagcccttcag
t2385c	ccaaaacgtgccagaactacgacctggagtgcat
t2385c_antisense	atgcactccaggtcgtagttctggcacgttttgg

Table 2-3: Primers used for pET01 plasmid preparation.

Variant	Sequence (5' to 3')
p.Arg924Gln_F	TTTCCCAGTCACGACGTTGGGATCCcagtggga ctctaatgggct
p.Arg924Gln_R	CAACGTCGTGACTGGGAAACGCCGGCGacac acacaaagaaacccgg
p.Leu1382=_F	TTTCCCAGTCACGACGTTGGGATCCtgtggatgg cctagaacaacgagt
p.Leu1382=_R	CAACGTCGTGACTGGGAAACGCCGGCGgggga agaggctgtgtgataaagt
VWF_exon_18_F	TTTCCCAGTCACGACGTTGGGATCCtcttttcaagcc tgggtctg
VWF_exon_18_R	CAACGTCGTGACTGGGAAACGCCGGCGgcccctg ggaaaattctagt

Table 2-4: Oligos used to prepare pcDNA-Dup (SF2-ASF3x) plasmid for ESE assay.

Variant	Sequence (5' to 3')
pcDupBR2int11F (-ve)	AATTCTGTTGTCCAGGTCACATTCAATTATAGGG
pcDupBR2int11R (-ve)	GATCCCTATAATTGAATGTGACCTGGACAACAG
2385T_F	AATTCAACGTGCCAGAACTATGACCTGGAGTGCATGG
2385T_R	GATCCCATGCACTCCAGGTCATAGTTCTGGCACGTTG
2385C_F	AATTCAACGTGCCAGAACTACGACCTGGAGTGCATGG
2385C_R	GATCCCATGCACTCCAGGTCGTAGTTCTGGCACGTTG
2365A_F	AATTCGAAGGGCTCGAGTGTACCAAAACGTGCCAGAG
2365A_R	GATCCTCTGGCACGTTTTGGTACACTCGAGCCCTTCG
2365G_F	AATTCGAAGGGCTCGAGTGTGCCAAAACGTGCCAGAG
2365G_R	GATCCTCTGGCACGTTTTGGCACACTCGAGCCCTTCG
2385T/2365A_F	AATTCAGTGTACCAAAACGTGCCAGAACTATGACCTG
2385T/2365A_R	GATCCAGGTCATAGTTCTGGCACGTTTTGGTACACTG
2385C/2365G_F	AATTCAGTGTGCCAAAACGTGCCAGAACTACGACCTG
2385C/2365G_R	GATCCAGGTCGTAGTTCTGGCACGTTTTGGCACACTG
5844C_F	AATTCGCGTGTGCACAGGCAGCTCCACTCGGCACAG
5844C_R	GATCCTGTGCCGAGTGGAGCTGCCTGTGCACACGCG
5844T_F	AATTCGTGTGTGCACAGGCAGCTCCACTCGGCACAG
5844T_R	GATCCTGTGCCGAGTGGAGCTGCCTGTGCACACACG
3789G_F	AATTCTGTGGAGGACATCTCGGAACCGCCGTTGCACG
3789G_R	GATCCGTGCAACGGCGGTTCCGAGATGTCCTCCACAG
3789A_F	AATTCTGTGGAGGACATCTCAGAACCGCCGTTGCACG
3789A_R	GATCCGTGCAACGGCGGTTCTGAGATGTCCTCCACAG
1451A_F	AATTCACCTCCGCATCCAGCATACAGTGACGGCCTCG
1451A_R	GATCCGAGGCCGTCACTGTATGCTGGATGCGGAGGTG
1451G_F	AATTCACCTCCGCATCCAGCGTACAGTGACGGCCTCG
1451G_R	GATCCGAGGCCGTCACTGTACGCTGGATGCGGAGGTG
2771G_F	AATTCTGAAATGCAAGAAACGGGTCACCATCCTGGTG
2771G_R	GATCCACCAGGATGGTGACCCGTTTCTTGCATTTCAG
2771A_F	AATTCTGAAATGCAAGAAACAGGTCACCATCCTGGTG
2771A_R	GATCCACCAGGATGGTGACCTGTTTCTTGCATTTCAG
2880G_F	AATTCAATTCGTGGAGTCTGGCCGGTACATCATTCTG CTGG
2880G_R	GATCCCAGCAGAATGATGTACCGGCCAGACTCCACG
2880A_F	AATTCGTGGAGTCTGGCCGATACATCATTCTGCTGG
2880A_R	R:GATCCCAGCAGAATGATGTATCGGCCAGACTCCACG
1548T_F	AATTCCTGTCCCCCGTCTATGCCGGGAAGACCTGCG
1548T_R	GATCCGCAGGTCTTCCCGGCATAGACGGGGGACAGG

1548C_F	AATTCCTGTCCCCCGTCTACGCCGGGAAGACCTGCG
1548C_R	GATCCGCAGGTCTTCCCGGCGTAGACGGGGGACAGG
4146G_F	AATTCCTCCCGCATCACCCTGCTCCTGATGGCCAGG
4146G_R	GATCCCTGGCCATCAGGAGCAGGGTCATGCGGGAGG
4146T_F	AATTCCTCCCGCATCACCCTTCTCCTGATGGCCAGG
4146T_R	GATCCCTGGCCATCAGGAGAAGGGTGATGCGGGAGG
4146A_F	AATTCCTCCCGCATCACCCTACTCCTGATGGCCAGG
4146A_R	GATCCCTGGCCATCAGGAGTAGGGTGATGCGGGAGG
4146C_F	AATTCCTCCCGCATCACCCTCCTCCTGATGGCCAGG
4146C_R	GATCCCTGGCCATCAGGAGGAGGGTGATGCGGGAGG

2.1.4 TaqMan® probes

TaqMan® probes (obtained from Life Technologies) were used for two different purposes in this project, which were, first, to study two SNV in *FUT6* (Table 2-5), and second, for quantification of mRNA for *VWF*, *B2M*, and *CLEC4M* (Table 2-6).

Table 2-5: TaqMan® probes used for genotyping *FUT6* variants.

Targeted SNV	Sequence (5' to 3')
c.370C>T rs778805	CCTCG[A[3]/G[4]]GGAGCGT[1]
c.977G>A rs61739552	ACGCTGC[G[3]/A[4]]GCCTC[2]

[1]TaqMan® assay probes previously designed and available (assay ID C_801652_20); [2]TaqMan® assay custom designed for the current study; [3]fluorescently tagged with VIC; [4]fluorescently tagged with 6-fluorescein amidite (6-FAM).

Table 2-6: TaqMan® probes used for mRNA quantification.

Target gene	Assay ID
VWF	Hs01109446_m1
B2M	Hs00984230_m1
CLEC4M	Hs03805885_g1

2.1.5 Plasmids

1.1.1.1 Splice analysis

The pcDNA-Dup (SF2-ASF3x) plasmid used for exon splice enhancer assays was generously donated by Dr. Pascaline Gaildrat (Inserm UMR 1079, Faculté de Médecine et de Pharmacie, Rouen, France) (Tournier*, et al* 2008, Gaildrat*, et al* 2012). The pET01 plasmid used for the acceptor / donor motif splicing assay was obtained from MoBiTec (GmbH, Germany) (Appendix 2).

1.1.1.2 VWF expression analysis

A bacterial plasmid expressing wild-type VWF was kindly provided by Prof. Reinhard Schneppenheim (Hamburg, Germany) (Schneppenheim*, et al* 2001). The plasmid was composed of full length *VWF* cDNA (8.6kb) inserted in a pcDNA3.1/Hygro (-) plasmid (Life Technologies) (5.6 kb); the total size of the plasmid was ~14.2kb (Appendix 3). This plasmid was used subsequently to investigate the effect of genetic changes on VWF expression in addition to the possible mechanism of any effect on expression level (for example RNA half-life).

1.1.1.3 Preparation of Renilla plasmid

A plasmid containing the Renilla gene (phRL-null) (Promega) was used in this study to normalise for transfection efficiency variation. This is a humanised version of the pRL-null plasmid, i.e. the codons have been replaced with more efficient human codons resulting in better expression in human cell lines (Promega.co.uk). The plasmid expresses renilla protein that emits florescence (from enzyme-independent oxidation) which is measured using luminometry. It was co-transfected along with VWF expression plasmid and measured independently to take into account variations that might occur during the transfection process such as variation in transfection efficiency or cell number. It was prepared using the maxiprep technique (section 2.2.2.1.11) from a previously prepared -80°C glycerol stock.

2.1.6 Cell line

The adherent Human Embryonic Kidney (HEK) 293T cell line, which was used for various experiments including RNA and protein quantification as well as splicing analysis, was obtained from LGC Standards (Middlesex, UK). HEK293T is an adherent cell line that mimics, to an extent, endothelial cells (by naturally producing pseudo Weibel-Palade bodies) but does not express endogenous VWF. In addition, these cells express SV40 large T antigen which binds to SV40 to enhance plasmid expression. Therefore, it was an excellent model to study VWF expression.

2.2 Methods

2.2.1 *In silico* analysis

In silico analysis was performed using several web-based tools. All tools were accessed March 2012 unless otherwise stated.

2.2.1.1 General *in silico* tools

Some web tools were used for general *in silico* purposes such as checking for the reference sequence of DNA using the University of California Santa Cruz (UCSC) Human Genome Browser (GRCh37/hq19) (http://genome.ucsc.edu/) and National Centre for Biotechnology Information (NCBI) (build number 37.2) (http://www.ncbi.nlm.nih.gov/) or for cDNA using the Ensembl Genome Browser (build 62) (http://www.ensembl.org/index.html). UCSC (PCR tool) was also used to give the expected product from a given primer pair. All three tools were also used to check for the position of any SNV, splice variants, exon/intron starting points, and surrounding genes. Finally, NCBI Basic local alignment search tool (BLAST) was further used to either align two given sequences or one given sequence with the human genome to identify similar sequences.

2.2.1.2 Protein prediction tools

Several web tools were used to predict the effect of a given variant on its protein function. The result from one web tool is not sufficient because each tool uses a different algorithm which investigates SNV from a different aspect (i.e. some SNV that have an effect might not be detected, or SNV that do not have an effect might be detected). Therefore several web tools were used in order to maximise the accuracy of the obtained results.

These tools include Sorting Intolerant from Tolerant (SIFT) http://sift.jcvi.org/, accessed April 2012, Align GVGD (Grantham Variation and Grantham Deviation) http://agvgd.iarc.fr/agvgd_input.php, accessed April 2012 (Mathe*, et al* 2006), polymorphism phenotyping V2 (PolyPhen-2) http://genetics.bwh.harvard.edu/pph2/, accessed April 2012 (Adzhubei*, et al*

2010) and mutation tasting http://www.mutationtaster.org/, accessed August 2015 (Schwarz*, et al* 2010).

2.2.1.3 mRNA prediction tools

Several web tools were used to predict the effect of given SNV on mRNA splicing for the same reason as using several protein prediction tools. mRNA prediction tools investigated two aspects of splicing. First, tools that investigate the effect of SNV on exon splice enhancers included ESEfinder2.0 http://rulai.cshl.edu/tools/ESE2/ (Cartegni*, et al* 2003) and Human Splicing Finder - Version 2.4.1 http://www.umd.be/HSF/ (Desmet*, et al* 2009).

Second, tools that investigated the effect of SNV on acceptor / donor splicing motifs. These tools include NetGene2 http://www.cbs.dtu.dk/services/NetGene2/ (Brunak*, et al* 1991), Splice Site Prediction by Neural networks http://www.fruitfly.org/seq_tools/splice.html (Reese*, et al* 1997), Automated Splice Site And Exon Definition Analyses (ASSEDA) http://splice.uwo.ca/ (accessed May 2014), Alternative Splice Site Predictor (ASSP) http://wangcomputing.com/assp/ (accessed May 2014) (Wang and Marin 2006) and Human Splicing Finder - Version 2.4.1 http://www.umd.be/HSF/ (Accessed May 2014).

In addition, RNAsnp web server http://rth.dk/resources/rnasnp was used to investigate the effect of a given SNV on mRNA secondary structure by comparing it to WT sequence (accessed May 2015) (Sabarinathan*, et al* 2013). miRNA was also investigated using miRBASE web tool http://www.mirbase.org/ (Accessed May 2015) (Griffiths-Jones*, et al* 2008, Kozomara and Griffiths-Jones 2011).

2.2.1.4 Other *in silico* tools

Other web tools includes Sequence Manipulation Suite (v2) http://www.bioinformatics.org/sms2/dna_stats.html which was used to calculate the GC percentage of a given sequence, ClustalW2 http://www.ebi.ac.uk/Tools/msa/clustalw2/ which was used to align given sequences (Larkin*, et al* 2007), dbSNP (build 132) http://www.ncbi.nlm.nih.gov/projects/SNP/ which was used to obtain

information on a given SNV such as rs number, a.a change and number, and nucleotide position and change as well as minor allele frequency in different populations, HapMap haploview (V4.2) http://hapmap.ncbi.nlm.nih.gov/ to investigate the linkage disequilibrium (LD) between different SNV (Barrett*, et al* 2005) and SNPStats (v0.95) http://bioinfo.iconcologia.net/SNVstats/start.htm, accessed July 2012, to calculate the LD between different SNVs from a given dataset and their association with a given response in addition to testing the Hardy Weinberg equilibrium (Sole*, et al* 2006). SnapGene viewer v.2.1 was used to view and manipulate plasmids (http://www.snapgene.com/, Accessed April 2014).

2.2.1.5 Oligo design and validation

2.2.1.5.1 Primers for PCR amplification

Primers were designed using two different methods. The Primer3 v0.4.0 web-tool (http://frodo.wi.mit.edu/primer3/, accessed March 2012) (Rozen and Skaletsky 2000) was used to design the majority of primers. Each primer was designed to meet several parameters to achieve the best results including primer length between 20-25 nucleotides, GC content between 45-55% and no single nucleotide to exceed four repeats.

Designed primers were validated using the UCSC genome browser PCR tool and NCBI BLAST in order to avoid non-specific binding of primers. The UCSC PCR tool indicated the expected product(s) of each primer pair, thus defining any possible nonspecific binding of the primers, while BLAST indicated whether similar nucleotide sequences to the designed primers occurred elsewhere in the human genome. Primer annealing sites were then checked for the presence of any SNV, which might completely prevent the primers from binding to DNA samples carrying the SNV (if homozygous for the SNV) or cause individuals heterozygous for the SNV to appear homozygous (due to the primer binding to one allele but not the other).

Primers were designed manually (following the same criteria) if Primer3 failed to design appropriate primers, if primers designed using Primer3 failed the validation process, or if primers needed designing in a specific location (for

example, internal sequencing primers for *FUT3*, *FUT5* and *FUT6* were designed at positions where all three genes were homologous).

2.2.1.5.2 Primers for site directed mutagenesis

Primers were designed for each specific mutagenic reaction using the provided tool at www.aglient.com/genomics/qcpd (accessed March 2013) according to the manufacturer's specifications: the desired change in the middle of the primer, primer size between 25-45 bp, melting temperature ≥ 78°C, GC content between 40-60% and the primers ending with C or G nucleotide.

For each mutagenesis reaction, primers were diluted to the appropriate concentration, equating to 125ng/µ calculated using the following equation

Required concentration (pmols/µl) = (desired ng X 1000) / (330 X length of primers).

2.2.1.6 Polymerase chain reaction (PCR)

PCR is a procedure that is used to produce identical and sufficient copies of a target DNA sequence to be used for subsequent investigation such as sequencing (Mullis 1990). PCR amplification goes through three main steps. The first step is denaturation, in which double stranded DNA is denatured (always at high temperature, separating double stranded DNA to allow the primers to bind). The second step is annealing, where primers attach to complementary target DNA. The last step is extension, where nucleotides bind next to the annealed primers, forming a complementary DNA strand.

The annealing temperatures (AT) were originally based on the primer melting temperature (T_M), calculated using the following formula:

$T_M = 69.3 + (GC\% \text{ X } 0.41) – (650 / \text{primer length})$

2.2.1.6.1 Standard PCR

The standard protocol was used to amplify small sized products (≤800bp). PCR amplification was performed using ReddyMix™ (Fisher Scientific UK Ltd., Loughborough) following the manufacturer's guidelines for the

preparation: 22.5 µl ReddyMix™ (containing 75mM Tris-HCl (pH 8.8), 20 mM $(NH_4)_2SO_4$, 0.01% Tween® 20, 1.5mM $MgCl_2$ and 0.2mM dNTPs), 1.0 µl dH_2O, 0.5 µl forward and reverse primers (at 10 pmol/µl) and 0.5 µl DNA (or dH_2O for a negative control) per 25µl reaction.

PCR amplifications were electrophoresed on a 96-Well GeneAmp® PCR System 9700 (Life Technologies, Warrington, UK), the conditions varying depending on the primer, the expected PCR product size and the protocol used. Standard PCRs were amplified using the following conditions: 94.0°C for 5 min, [94.0°C for 30 sec, X °C (annealing temperature which is primer specific) for 30 sec, 72.0°C for 30 sec] for 35 cycles, 72.0°C for 10 min, then cooling to 4°C until being removed from the thermocycler.

2.2.1.6.2 Long-range PCR

A long range PCR protocol was used to amplify products of larger sizes (>800bp) and was performed using OneTaq® Hot Start DNA polymerase (New England Biolabs). The manufacturer's guidelines were followed for the preparation: 12.5µl OneTaq® Hot Start Master Mix (containing 20 mM Tris-HCl (pH8.8), 22 mM KCl, 0.2 mM dNTPs, 1.8 mM $MgCl_2$, 5% glycerol, 0.05% Tween®-20, 25 units/ml OneTaq DNA Polymerase, 22 mM NH_4Cl, and 0.06% IGEPAL® CA-630), 10.5µl dH_2O, 0.5µl forward and reverse primers (at 10pmol/µl), and 1.0µl DNA (or dH_2O for a negative control) per 25µl reaction.

Long range PCRs were amplified using the following conditions: 94.0°C for 30 sec, [94.0°C for 30 sec, 57.0°C for 45 sec, 68.0°C for 70 sec] for 30 cycles, 68.0°C for 5 min.

2.2.1.6.3 High Fidelity PCR

High Fidelity polymerases were used to amplify products with much more amplification (copying) accuracy, specifically when amplified products were to be used for cloning purposes. Two high-fidelity polymerases were used during this project, KAPA HiFi and Q5 high-fidelity. For both polymerases, the manufacturer's guidelines were used. For KAPPA HiFi, 10µl 5X KAPA HiFi Fidelity Buffer, 1.5 µl 10mM KAPA dNTPs Mix, 1.5 µl 10 µM forward primer, 1.5 µl 10 µM reverse primer, 1.0 µl 100ng DNA in 50 µl reaction made up with

dH_2O. For Q5 high-fidelity, 5µl 5X Q5 reaction buffer, 0.5 µl 10mM dNTPs, 1.25 µl 10mM forward primer, 1.25 µl 10mM reverse primer, 1.0 µl 100ng DNA, 0.25 µl Q5 high-fidelity DNA polymerase made up to 25 µl with dH_2O. PCR was performed using the following conditions: KAPA HiFi, 95.0 °C for 3 min, [98.0 °C for 20 sec, X°C for 15 sec, 72.0 °C for 40 sec] for 35 cycles, 72.0 °C for 1 min; Q5 high-fidelity, 98.0 °C for 30 sec, [98.0 °C for 10 sec, X°C for 30 sec, 72.0 °C for 30 sec] for 30 cycles, 72.0 °C for 2 min.

2.2.1.6.4 PCR troubleshooting

If the initial PCR failed to work, the temperature was adjusted either by increasing (to increase specificity) or decreasing (if no products were obtained) it slightly. On the other hand, the magnesium concentration in each kit already provides the reaction with the required amount of magnesium. However, the magnesium concentration can be increased, which in turn increases the activity of the polymerase and reduces the specificity for greater chance of primer annealing.

2.2.1.7 Agarose gel electrophoresis

In order to check whether the PCR amplification had succeeded and that products of the expected size had been obtained, agarose gel electrophoresis was performed. The idea behind gel electrophoresis is that an electrical charge is run through the gel migrating the loaded DNA (which is positively charged) from the top (positive charge) to the bottom (negative charge) of the gel. DNA with different sizes move at different speeds through the gel pores (i.e. larger DNA fragments move more slowly than smaller fragments), resulting in the separation of the loaded DNA according to size.

Different agarose gel concentrations were used depending on the size of the PCR product; products ≤700bp were run in 2% w/v agarose gels and products >700bp were run in 1% agarose gels. Decreasing the concentration of the agarose increased the pore size within the agarose gel allowing larger products to migrate faster through the gel.

Gels were made using 1 or 2g agarose (Bioline) dissolved in 100ml 0.5x Tris-borate ethylenediaminetetraacetic acid (EDTA) buffer (1X TBE containing:

0.89M Tris, 0.89M Boric Acid and 0.02M EDTA) with a final concentration of 1mM ethidium bromide (3,8-diamino-5ethyl-6 phenylphennanthridinium bromide). Ethidium bromide interacts with DNA and emits orange fluorescence that becomes stronger after binding to DNA and makes it visible under UV lighting. For each gel, 5µl of sample DNA was loaded in each lane and run against a HyperLadder™ size standard (Bioline) added to another lane. Gels were run in 0.5x TBE at 80-130V (voltage reduced for smaller sized products to allow better separation) for 50 min and visualised using a UV transilluminator (Bio-Rad Laboratories Ltd., Hemel Hempstead, UK).

2.2.1.8 Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis was used to check for small size oligos (less than 100bp). 4% Polyacrylamide gel was prepared as follows: 1.88ml 40% acrylamide/Bis (w/v), 1.50ml 5x TBE and 11.63ml dH_2O were mixed. Then 150µl 10% ammonium persulfate (APS) and 15µl Tetramethylethylenediamine (TEMED) were added to begin polymerisation. The gel was immediately poured into a prepared cassette and was left to completely solidify for 1h. 5µl sample or DNA ladder was added to each well. The gel was run in 0.5x TBE at 200V, 150mA for ~30mins. The gel was gently removed and incubated in 100ml 0.5x TBE with 10µl ethidium bromide for 15min on an orbital shaker at 20rpm before being visualised using a UV transilluminator (Bio-Rad).

2.2.1.9 PCR purification

PCR purification was performed to remove any unwanted components within amplified products which might disrupt downstream applications, for example DNA sequencing. The ExoSAP-IT® reagent (GE Healthcare, Little Chalfont, UK) was primarily used in this study. ExoSAP-IT® reagent is based on Shrimp Alkaline Phosphatase and Endonuclease I; it degrades unused primers following PCR and dephosphorylates dNTPs for better and cleaner reading. The ExoSAP-IT® purification method was performed using 6µl reagent added to 6µl PCR product and incubated at 37.0 °C for 30 min (to activate the enzymes), 80.0 °C for 10 min (to degrade the enzymes), and 14.0 °C for 7 min.

Some products were not adequately purified using ExoSAP-IT®, so the AMPure XP purification system (Beckman Coulter (UK) Ltd., High Wycombe) was used instead (performed by the Core Genetics Facility, Medical School, University of Sheffield). This method uses magnetic beads to bind to the DNA products, washes are subsequently applied to eliminate any contaminants unbound to the magnetic beads resulting in clean DNA products.

Gel extractions were also used for DNA purification. If there were non-specific products and a band(s) of the expected size that needed to be selectively extracted or after performing restriction enzyme digestion and unwanted digested products needed eliminating, then gel extractions were performed using a QIAquick Gel Extraction Kit (QIAGEN Ltd.) following the manufacturer's guidelines. The gel was cut (around the product of interest) using a sterile scalpel. It was then weighed and QG buffer was added (to dissolve the gel) at a volume equal to 3X the gel weight (in g). The gel was then incubated at 50°C until completely dissolved. Isopropanol was added at a volume equal to 1X the gel weight and the mixture was transferred to the provided QIAquick column. The column was centrifuged at 13000 x g for 1 min, 500µl QG buffer was added and the column centrifuged again at 13000 x g for 1 min. Next, 750µl PE buffer (wash buffer) was added, the column centrifuged at 13000 x g for 1 min, the flow through was discarded and another minute of centrifugation at the same speed was performed. Finally, the column was transferred to a 1.5mL micro-centrifuge tube, 20µl elution buffer (EB) buffer was added to the column and incubated at room temperature for 1 min. Following incubation, the column was centrifuged at 13000 x g for 1 min to elute the sample.

2.2.1.10 DNA sequencing and analysis

DNA sequencing was performed by the Core Genetics Facility using the Big Dye® Terminator v3.1 cycle sequencing kit (Applied Biosystems). Sequenced products were cleaned (i.e. had impurities which could reduce the sequencing quality removed) using the Agencourt® CleanSEQ kit (Beckman Coulter) and run on an ABI 3730 DNA analyzer (Life Technologies). The principle of sequencing is that DNA denaturation occurs at high temperature where the

double-stranded DNA is separated. Following that, annealing of a primer takes place and nucleotides (labelled with different fluorescent dyes) bind according to the sequence; when a nucleotide is attached it releases a signal (pre-attached dye) that is then detected and the assigned nucleotide is reported.

Generated sequence data was analysed using the Staden package v4.8b1 (Bonfield*, et al* 1998), which allows generated sequences to be compared directly against a known reference sequence allowing identification of variants (Section 2.2.1.1).

Following mutagenesis, the sequence data provided in FASTA format was compared to the reference sequence (*VWF* Gene ID: ENSG00000110799) using the NCBI BLAST tool to check for the presence of the desired change. In addition, the rest of the *VWF* cDNA was also checked for the presence of any additional changes that might have accidently been created during the mutagenesis process.

2.2.1.11 Genotyping

2.2.1.11.1 TaqMan genotyping

TaqMan® analysis uses quantitative real-time PCR (qRT-PCR) to investigate the presence of a specific nucleotide allele using a pair of fluorescently tagged probes each specific for a different allele. When either probe attaches to the target DNA, amplification releases a specific fluorescent dye (for that specific allele) which is then detected (Figure 2-1). In this study, TaqMan® was used to analyse two SNVs in *FUT6* (c.370C>T and c.977G>A) and was performed by the Core Genetics Facility using an Applied Biosystems® 7900HT Fast Real-Time PCR System (Life Technologies).

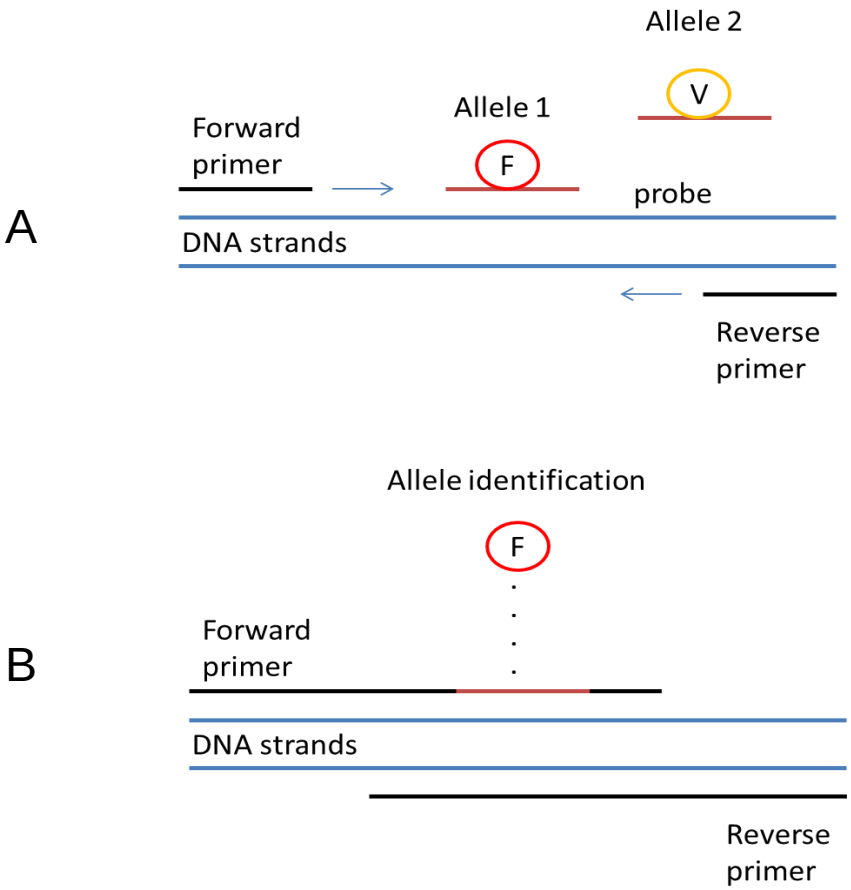


Figure 2-1: Principle of the TaqMan® genotyping method. Primers and specific probes first attach to the target DNA **(A)**. When amplification takes place, allele identification occurs after the dye is released **(B)**. Each dye is specific for one allele, thus, depending on the released dye(s), the related allele is called. F= FAM dye while V=VIC dye.

2.2.1.11.2 VNTR genotyping

VNTR genotyping was performed to accurately measure the size of amplified PCR products. Fluorescently labelled primers were used to amplify the region of interest (and 15bp of flanking DNA sequence) and PCR products were run on an ABI 3730 DNA Analyzer (Life Technologies; performed by the Core Genetics Facility). PCR products were run against a GeneScan™ 1200 LIZ® size standard (Life Technologies) and size analysis was performed using Peak Scanner™ software v1.0 (Life Technologies) (Figure 2-2).

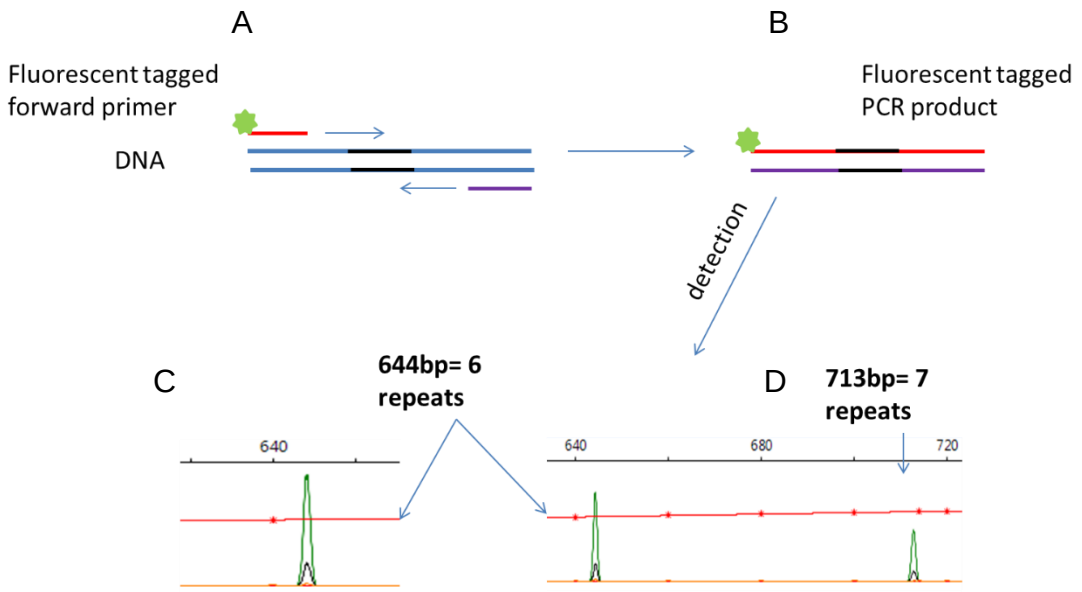


Figure 2-2: Principle of the VNTR Genotyping method. Amplification of the target DNA is performed using a fluorescently (6-JOE, SE (6-Carboxy-4',5'-Dichloro-2',7'-Dimethoxyfluorescein, Succinimidyl Ester) tagged forward primer **(A)**. This generates a fluorescently tagged PCR product **(B)**. The size of the PCR product(s) indicates the genotype, for example an individual homozygous for 6 repeats, **(C)** or an individual heterozygous for 6 and 7 repeats **(D)**.

2.2.2 *In vitro* analysis

In order to validate the findings obtained using *in silico* analysis and genotyping, further experiments were performed *in vitro* using HEK293T cells.

2.2.2.1 Tissue culture

HEK293T cells were grown in Dulbecco's modified Eagle's medium (DMEM supplemented with GlutaMAX™) (Life Technologies) to which 10% v/v foetal bovine serum (FBS) (Life Technologies) was added. GlutaMAX™ (Life Technologies) helps increase the stability of the media, increase cell growth performance and reduce ammonia toxicity. All tissue culture work was performed in a sterile tissue culture hood and all reagents used were pre-warmed to 37°C.

2.2.2.1.1 Cell thawing

Thawing refers to removing stored cells from liquid nitrogen and initiating their growth. Removed cells were defrosted using a 37°C pre-heated water bath until the storage media was 90% defrosted to avoid over heating the cells. The stored cells were then mixed with 5ml media and centrifuged at 1500 x g for 5 min. The media was then discarded and the cell pellet re-suspended in 13ml fresh media and incubated in a T75 flask at 37°C and 5% CO_2 in a humidified environment until 80-90% confluent.

2.2.2.1.2 Cell passaging (splitting)

When cells reached 80-90% confluence, they became ready for passaging. Although the colour of the culture media could be used as an indication on the readiness of the cells for passaging (fresh media was red in colour and began turning yellow when cells started to become confluent, resulting from a change in pH), a microscope was also used to check the confluence of the cells and percentage confluence was recorded depending on the percentage of area covered with cells. A rapid (1 day) change in colour from red to yellow was also used as an indication of infection and if this occurred the cells were investigated under the microscope and disposed of if contamination observed.

Once cells were ready for passaging, the media was discarded and the cells washed with 7ml phosphate-buffered saline (PBS) (Life Technologies) to remove any leftover media and dead cells. 2ml 1X trypsin prepared by adding 10ml 5X trypsin to 40ml Dulbecco's PBS (DPBS, Life technologies) was then added to detach the cells from the flask. The cells were incubated for 3 min followed by addition of 7ml media (to inhibit the trypsin activity) and centrifuged at 1500 x g for 5 min. The media was then discarded and the cell pellet re-suspended in 5ml fresh media. Depending on the pellet size 300-500µl of cell suspension was then added to a new T75 flask with 13ml fresh media.

2.2.2.1.3 Cell counting

Cells were counted when a specific number of cells was required to perform a specific downstream experiment. The protocol for cell passaging was followed until the cell suspension was prepared. At this stage 20µl of cell suspension was then added to a counting chamber (AC6000 Modified Fuchs Rosenthal haemocytometer, Hawksley, Sussex, UK), after which the cells were then counted in four 4X4 small squares under a microscope at 10X magnification (Figure 2-3). The total number of cells was then divided by 4 to obtain the average number of cells which was then multiplied by 10 to get the number of cells per 1µl (as each square contained ~0.1µl media).

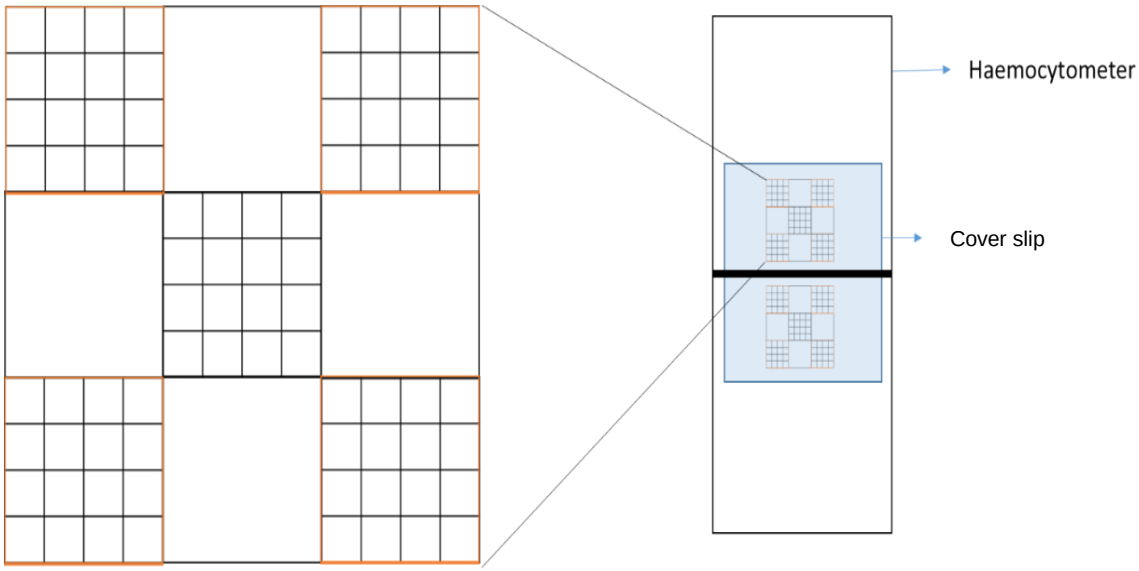


Figure 2-3: Cell counting method. Four 4X4 squares were counted (orange boundary) and the results divided by four to obtain the average cells per square which was then multiplied by 10 to indicate the total cells in 1 µl of cell suspension.

2.2.2.1.4 Freezing of cells

Cells were frozen down at early passages (2nd to 3rd passage) and stored in a -80°C freezer (for up to one year) or liquid nitrogen (for long term storage). Freezing was performed by mixing $1X10^6$ cells with 1ml of freezing media (70% DMEM, 20% FBS and 10% dimethyl sulfoxide (DMSO) v/v).

2.2.2.2 DNA preparation

2.2.2.2.1 Site-directed mutagenesis

Mutagenesis was performed on the VWF plasmid to study the effect of different genetic changes on VWF expression using a QuikChange® Lightning site-directed mutagenesis kit (Agilent Technologies).

Mutagenesis was performed in 50µl reactions which included 5µl reaction buffer, 125ng each primer, 1µl dNTP mix, 1.5µl Quick solution reagent, 100ng/µl plasmid (WT VWF expression plasmid), and 1µl QuickChange lightening enzyme. Finally, dH_2O was added to bring the reaction mix to 50µl.

Reactions were run on a thermal-cycler using the following settings: 95°C for 2 min, [95°C for 20 sec, 60°C for 10 sec, 68°C for 7 min and 30 sec] for 18 cycles, 68°C for 5 min. Next, the reaction was treated with 2µl *Dpn*I restriction enzyme and incubated at 37°C for 15 min to digest the original (methylated) plasmid. The product was then used to transform a competent *E.coli*.

2.2.2.2.2 Transformation

Transformation refers to insertion of plasmid DNA of interest into competent *E.coli* cells which in turn divide to further produce a large number of copies of the inserted plasmid. Two different competent strains were used in this project, NM554 (previously prepared, section 2.2.2.1.7) and XL-10 gold ultracompetent cells (provided with the mutagenesis kit). The *Dpn*I treated mutagenesis product was transformed alongside dH_2O (to check for possible contamination) and previously prepared wild-type plasmid DNA, which was used as a positive control. Slightly different transformation protocols were used for each of the two competent strains.

In the case of XL-10 gold ultracompetent cells, the cells were gently defrosted on ice and 45µl were transferred to a 14-ml BD Falcon polypropylene round-bottom tube to which 2µl ß-mecraptoethanol (ß-ME) was added to increase the transformation efficiency. 9µl of *Dpn*I treated product was then added to the cells and mixed well, and the whole mixture incubated on ice for 15 min. Following incubation, heat shock at 42°C for 45 sec was carried out and then the mixture was transferred to ice and incubated for a further 5 min. 250µl preheated (at 42°C) NZY broth was then added and the whole mixture incubated at 37°C for 45min on an orbital shaker. Finally, the whole mixture was plated on a prepared agar plate with 0.1% ampicillin (1M sodium bicarbonate and 1.2M of ampicillin trihydrate (Sigma)) and incubated overnight at 37°C. Ampicillin is a selective reagent, which means it eliminates contamination and prevents the growth of competent cells without the plasmid of interest because the plasmid contains an ampicillin resistance gene.

Using NM554 competent cells required some alteration to the XL-10 protocol. 90µl of gently defrosted cells were transferred to a 14-ml BD Falcon polypropylene round-bottom tube (no ß-ME was added). 14µl of *Dpn*I treated product was then added and heat shock performed at 42°C for 90 sec. 150µl preheated (at 42°C) LB broth was added per reaction. The mixture was then plated as described above.

The next day, the plates were checked for availability of colonies, if colonies were available the plates were moved to the fridge to stop further colony growth, whilist, if no colonies were available, the plates were incubated for an additional 6 h.

2.2.2.2.3 Preparation of competent *E.coli* NM554 cells

Competent *E.coli* are bacterial cells that can incorporate foreign DNA after treatment of their cell membrane, allowing DNA to pass through. A scraping of previously prepared NM554 glycerol stock was grown in 10ml LB broth overnight at 37°C in an orbital shaker. The culture was then transferred to 500ml LB broth and grown for an additional 2 h at 37°C in the orbital shaker. The culture optical density (OD) was measured using a Biophotometer

spectrophotometer (Eppendorf UK Ltd., Stevenage) at 600nm. The culture was used when an OD between 0.2-0.4 was achieved, indicating that the cells were in log phase. The culture was then centrifuged at 6400 x g for 7 min and the LB broth discarded, after which the cell pellet was re-suspended in 40ml transformation buffer (prepared using 15% glycerol, 10mM PIPES (piperazine-N,N′-bis 2-ethanesulfonic acid) and 50mM $CaCl_2$ w/v). The mixture was then incubated on ice for 20 min followed by centrifuging at 6400 x g for 7 min. The cell pellet was then re-suspended in 25ml transformation buffer. Aliquots of 300µl were prepared and stored at -80°C. To assess the competent cells, negative control (dH_2O only to check for possible contamination) and positive control (VWF WT expression plasmid to test that competent cells were working as intended) transformations were performed and colony growth was observed.

2.2.2.2.4 Bacterial plasmid miniprep

Bacterial miniprep was performed to extract plasmid DNA from the grown competent cells. It was used to extract a low yield of transformed plasmid for further testing before performing bacterial maxiprep which is intended to produce a much higher yield.

Bacterial colonies were inoculated overnight in 8ml LB broth with 0.1% ampicillin at 37°C on an orbital shaker. 2ml was used to prepare a glycerol stock (Section 2.2.2.1.9) and the rest used for DNA extraction utilising the QIAprep Spin mini-prep kit (QIAGEN Ltd.) following the manufacturer's guidelines as follows. Cultured broth was centrifuged at 5000 x g for 5 min, then the broth was discarded and the pellet was suspended in 250µl buffer P1 (re-suspension buffer). 250µl buffer P2 (lysis buffer) to lyse the cells was added and the mixture was incubated for 5 min at room temperature. Stopping the lysis reaction was performed by adding 350µl N3 buffer. Following that, the mixture was centrifuged at 13000 x g for 10 min to precipitate the cell contents. The supernatant was then transferred to the provided filter and centrifuged at 13000 x g for 1 min to allow the DNA to bind to the filter. After that, washing the cell residue was performed using 750µl PE buffer (washing buffer) and centrifuged at 13000 x g for 1 min. 30-50µl (depending on the cell

pellet size) EB buffer (elution buffer) was added to the filter and incubated at room temperature for 1 min and then centrifuged at 13000 x g for 1 min. The DNA concentration was then measured (Section 2.2.2.1.10).

2.2.2.2.5 Glycerol stock

The purpose of making a glycerol stock was to create a stock of transformed *E.coli* for long-term storage at -80°C. It was performed by centrifuging 2ml pre-inoculated LB broth at 5000 x g for 5 min, discarding 1ml and re-suspending the pellet in the remaining 1ml LB broth to concentrate the bacterial stock. Following that, 300µl 10% glycerol was added. Glycerol is a cryoprotectant substance which protects the cells from freezing damage.

2.2.2.2.6 DNA concentration

DNA concentration was measured using a NanoDrop® ND-1000 (Thermo Fisher Scientific, Inc., Wilmington MA, USA) at an absorbance of 260nm. 1µl of dH_2O or EB was used as a blank, depending on the solution in which the plasmid DNA had been re-suspended. Following that, 1µl of each sample was used to measure the concentration (ensuring no air bubbles were present and that the DNA was well mixed for accurate measurement). The mean of three measurements was used. The quality of the DNA was also checked using the given ratio of A260/A280, with results from 1.8-2.3 considered of good quality, while higher or lower values would indicate protein contamination. In addition, the ratio of A260/A230 was also checked, with a ratio >2 considered of good quality; whereas <2 indicated the presence of contaminating salt impurities.

2.2.2.2.7 Bacterial plasmid maxiprep

A scraping of prepared -80°C glycerol stock was taken and incubated in 250ml freshly prepared LB broth with 0.1% ampicillin overnight at 37°C on the orbital shaker. The broth was then transferred to a clean 250ml Beckman tube and centrifuged at 6000 x g for 15 min at 4°C. The broth was discarded and the pellet re-suspended in 10ml buffer P1 (re-suspension buffer) followed by the addition of 10ml buffer P2 (lysis buffer) and incubation for 5 min at room temperature. Next, 15ml buffer P3 (neutralising buffer) was added to stop the lysis process. The mixture was poured into a capped cartridge and incubated

for 5 min at room temperature to allow the solid cell contents to separate from the liquid part containing the DNA. At this point 10ml QBT equilibration buffer (which enhances the DNA binding activity of the filter) was also added to a QIAGEN-tip 500 filter and was allowed to flow through using gravity. The cartridge was then uncapped and the solution was forced through the cartridge into the pre-equilibrated filter and allowed to flow through using gravity. The filter was then washed twice using 30ml of QC washing buffer (which cleans cell residue from the filter). 10ml QF elution buffer was then added to the filter and allowed to flow through using gravity into a clean 40ml Beckman tube. 10.5ml isopropanol was then added (to precipitate the DNA) and the mixture centrifuged at 15000 x g for 30 min at 4°C. Following that, the solution was discarded carefully and the DNA pellet washed using 5ml 70% ethanol. The solution was centrifuged at 15000 x g for 15 min at 4°C. The supernatant was discarded carefully and the tube was air-dried at room temperature for 10 min. Finally, the DNA pellet was re-suspended using 300-500µl of dH_2O depending on the pellet size. The DNA was stored at -20°C.

2.2.2.3 Protein and mRNA quantification

2.2.2.3.1 Transfections

The aim of transfection is to get the required plasmid into the nucleus of the cell to be expressed in order to study the product. There are two types of transfections which are transient and stable transfection. Transient transfection is used for short term study where the plasmid DNA does not integrate with the host cell DNA, while stable transfection is used for long term study and the plasmid DNA integrates with the host cell DNA; transient transfection was used in this study. Transfection was used in this project to express VWF in HEK293T cells and study the effect of different variants on protein expression, and RNA expression and half-life.

Before proceeding with transfection, 250,000 HEK293T cells were seeded into each well of a 6-well plate (Section 2.2.2.1.2). The seeded cells were incubated for 24 h prior to transfection to allow the cells to attach to the plate surface. Following that, transient transfection was carried out using transfection reagent Xfect (Takara Clontech, Saint-Germain-en-Laye France)

as follows. Per well, a mixture of 100µl Xfect buffer, 2.5µg expression plasmid DNA, 1.0µg Renilla plasmid and 1.1µl Xfect was prepared, mixed well and incubated at room temperature for 10 min. Then the mixture was added to each well and incubated for 24 h. Following that, the media was discarded and 2ml fresh media was added (due to the toxicity of the transfection reagent) to each well and the plate was incubated for an additional 48 h before harvesting.

For each variant expressing plasmid, three transfection experiments were performed independently each of which was carried out using three wells using either 100% WT VWF expression plasmid (to mimic homozygous WT state), 100% mutant (variant of interest expressing plasmid to mimic homozygous mutant state) and 50:50 co-transfection of WT and mutant plasmids (to mimic the heterozygous state of that variant).

After final incubation, harvesting took place in the exact same way regardless of transfection reagent used as follows: the culture media was collected in separate clean tubes and the wells washed using 400µl PBS. Following that, 300µl of 1X passive lysis buffer (Promega) was added to the wells (to lyse the cells and extract VWF from inside the cells) and the plate incubated at room temperature for 10 min at 60 rpm. The media and cell lysate were then stored at -20°C until needed.

Transfection for mRNA quantification was conducted using Xfect as follows. 250,000 HEK293T cells were seeded as per Section 2.2.2.1.2 in 6-well plates and incubated for 24 h. Following that, a mixture of 100µl Xfect buffer, 0.9µl Xfect and 2.5µg DNA was incubated at room temperature for 10 min followed by the addition of the mixture to each well. The wells were incubated for 24 h post transfection. Following that, media was discarded, wells were washed with 0.5X PBS followed by the addition of 300µl trypsin and incubation of 5 min at room temperature was carried out. Next, 1ml of fresh media was added and the mixture was transferred to a 1.5ml Eppendorf tube. The tubes were centrifuged at 4000 x g for 5 min, the media was discarded and the cell pellet was used for RNA extraction. The experiment was repeated three independent times.

2.2.2.3.2 Protein quantification

2.2.2.3.2.1 Enzyme-linked immunosorbent assay (ELISA)

ELISA is a technique used to measure protein level by comparing against a standard curve of the protein of interest at known concentrations. Several different ELISA methods are available, however, in this project, the sandwich based ELISA method was used. It uses a primary antibody that binds to the surface of the well, followed by the addition of the sample where the target protein binds to the primary antibody. Next, a secondary antibody is added which binds to the protein that is bound to the primary antibody, substrate is then added which can be detected and quantified (Figure 2-4).

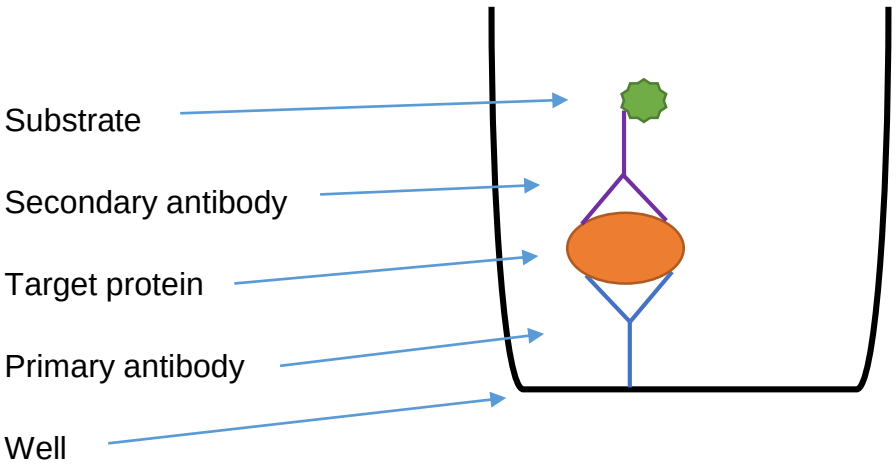


Figure 2-4: ELISA sandwich assay system.

For measurement of VWF, the antibodies (obtained from Enzyme Research) were used to perform ELISA according to the manufacturer's guidelines. The primary (capture) antibody was diluted in coating buffer (50 mM carbonate coating buffer, pH 9.6) 1 in 100 and 100µl added to each well and incubated overnight at 4°C to allow it to bind to the 96-well (flat bottom) plate. The antibody was discarded and 150µl blocking buffer (PBS-BSA, 1% w/v, pH7.4) added and incubated for 1 h at room temperature to block unbound areas and prevent further non-specific binding. Following that, the wells were washed twice using 100µl washing buffer (PBS-Tween, 0.1% v/v) to remove any unbound antibodies. 100µl of collected cell lysate or supernatant was then added and incubated for 90 min at room temperature to allow VWF to bind to the primary antibody. Two washes were then performed (to remove unbound sample) using 100µl washing buffer per well. Next, 100µl of 1/100 secondary (detection) antibody in sample diluent (100mM HEPES, 1% w/v BSA, 0.1% v/v Tween-20, 100mM NaCl, pH7.2) was added to bind to the bound VWF. Finally, two washes were carried out (to remove unbound secondary antibody) followed by the addition of 100µl of o-phenylenediamine (OPD) per well (prepared with one OPD tablet (Sigma Aldrich, UK) dissolved in 25ml substrate (27mM Citric acid, 97mM Na2HPO4, pH5.0), and 50µlH_2O_2). The plate was incubated for 5-10 min to allow the colour to develop. The reaction was stopped by adding 50µl of 2.5M sulphuric acid and the plates were then read using the Varioskan Flash plate reader at 490nm (Thermo Fisher Scientific).

The required standard curve was also prepared on the same reaction plate using a standard calibrator plasma (Quadratech). Eleven different dilutions (in sample diluent) of calibrator plasma were used: 1/10, 1/20, 1/40, 1/80, 1/160, 1/320, 1/640, 1/1280, 1/2560, 1/5120, 1/10240, with the last well left without calibrator. The calculations of the standard curve values were obtained using a previously designed Excel sheet (obtained from Nasher Alyami, Haemostasis Research Group).

Optical density (OD) data obtained from the plate reader was converted using the VWF standard curve of known concentrations into VWF:Ag levels, using GraphPad Prism v5.04 (GraphPad Software Inc., La Jolla, CA, USA).

2.2.2.3.2.2 Normalisation and analysis

Renilla luminescence was measured in cell lysates using the Renilla Luciferase Assay System (Promega) according to the manufacturer's guidelines. 1µl 100X Renilla assay substrate was diluted in 100µl Renilla assay buffer and added to 20 µl cell lysate. The 94 well plates were read immediately at 480nm.

Following that, the analysis was performed by normalising ELISA values using the Renilla readings to obtain a more accurate measurement of VWF concentration using Excel software. The final values of 100% mutant and 50:50 co-transfections were then compared to WT transfections (considered as 100%) to measure the effect of VWF mutations on VWF:Ag levels.

2.2.2.3.3 mRNA Quantification and mRNA half-life

Measurement of mRNA levels was carried out in this project to assess the effect of different genetic changes on RNA expression level.

2.2.2.3.3.1 RNA extraction

RNA extraction was performed using an EZ-RNA total RNA isolation kit (Geneflow Ltd., Lichfield, UK) following the manufacturer's guidelines. Previously prepared cell pellets were re-suspended in 500µl reagent A (re-suspension buffer) and incubated at room temperature for 5 min. Following that, 500µl reagent B (extraction buffer to break down the cells and extract the RNA) was added and incubated at room temperature for 10 min. The mixture was then centrifuged at 12000 x g for 15 min. The top clear layer of the resulting supernatant (which contains the RNA) was transferred to a new tube and the pink layer (which contains cell debris) was discarded. 500µl isopropanol was then added (to precipitate the RNA) and incubated at room temperature for 10 min followed by centrifuging the mixture at 12000 x g for 8 min. The mixture was then stored at -20°C overnight to increase the yield of the RNA. Next, the mixture was centrifuged again at 12000 x g for 8 min and the supernatant discarded. The RNA pellet was washed using 1ml 75% ethanol and centrifuged at 13000 x g for 15 min. The supernatant was then

discarded and the pellet air dried at room temperature for 5 min. Finally, the pellet was re-suspended in 30-50µl 0.1% diethylpyrocarbonate (DEPC) – treated water (RNAse free water) and stored at -20°C.

2.2.2.3.3.2 Reverse transcription

Reverse transcription was carried out to produce the complementary DNA (cDNA) required for quantification using quantitative real time polymerase chain reaction (qRT-PCR). A QuantiTect Reverse Transcription Kit (QIAGEN) was used for this purpose following the manufacturer's guidelines. 1µg RNA was dissolved in 10µl sterile dH_2O followed by the addition of 2µl genomic DNA wipe-out solution. The mixture was then incubated at 42°C for 10 min before being transferred to ice. Following that, 1µl reverse transcriptase, 1µl provided primer mix and 4µl reverse transcription buffer were added and the whole mixture incubated at 42°C for 30 min and then at 95°C for 3 min to de-activate the enzyme. Finally, the cDNA was stored at -20°C. The experiment was performed on ice due to the sensitivity of RNA to heat and using a sterile environment to avoid any possible DNA contamination.

2.2.2.3.3.3 qRT-PCR

qRT-PCR was performed to quantify the RNA in a given sample following its conversion into cDNA. This procedure uses a probe (with fluorescent dye) made specifically for the specific target sequence in addition to forward and reverse primers covering the region where the probe is expected to bind. The primers are prepared to bind to exon-exon boundaries to avoid binding and detecting DNA. Each cycle starts with cDNA denaturation, followed by annealing of primers and probe, amplification and finally detection. Following each cycle, after the amplification takes place any dye released is detected, measured and reported at the end of each cycle. A relative quantification method was used to measure the RNA level. This method relies on comparing the measurement of the gene of interest in relation to the level of an endogenous control.

TaqMan® Gene Expression Master Mix was used following the manufacturer's guidelines. Duplex qRT-PCR was used which measures the target (*VWF*;

assay ID Hs01109446_m1) and endogenous control (Beta-2 microglobulin (*B2M*); assay ID Hs00984230_m1) in the same well. This had the benefit of reducing the variation between *VWF* and *B2M* readings compared to taking readings in two separate wells (Livak and Schmittgen 2001).

100ng cDNA, 1µl *VWF* probe, 1µl *B2M* probe and 16µl master mix were mixed and added per sample per well in a 384-well plate. The plate was centrifuged at 2000 x g for 1 min to eliminate air bubbles and analysed on an Applied Biosystems 7900HT Fast Real-Time PCR System (Life Technologies) using the following conditions. 2 min at 50°C, initial denaturing for 10 min at 95°C followed by 50 cycles of denaturing for 15 sec at 95°C and extension for 1 min at 60°C.

After completion of the analysis, ΔΔ cycle threshold (CT), which is the difference of expression between two genes, was measured (by deducting the values of the endogenous control (i.e. *B2M*) from the values of the target (i.e. *VWF*)) (Livak and Schmittgen 2001). The final value was taken as an indication of VWF RNA quantity.

2.2.2.3.3.4 Measurement of mRNA half-life

The purpose of measuring mRNA half-life was to test the effect of different variants on mRNA stability. mRNA half-life was measured using the same qRT-PCR protocol (Section 2.2.2.1.12.3) except that 24 h after transfection, the cells were treated with 5µg/ml actinomycin D (Life Technologies), which is a drug that inhibits the transcription process (through binding to DNA at the transcription initiation complex and preventing RNA formation), thus inhibiting the production of more mRNA. Cells were harvested at 0, 2, 3 and 4 h post-treatment and half-life was calculated as the period of time in which mRNA level was 50% of the original level detected at 0 h post treatment (Leclerc*, et al* 2002).

2.2.2.4 RNA splicing assays

2.2.2.4.1 Exon splice enhancer (ESE) assay

An ESE assay was used to investigate the effect of given SNV on exon splice enhancers or silencers using a pcDNA-Dup (SF2-ASF3x) plasmid. For splicing

to take place, it requires splicing site in addition to splice enhancer sites. Therefore, loss of splicing enhancer site(s) or gain of splice silencer site(s) could result in reduction of splicing efficiency or even complete loss of splicing.

The plasmid used has three expressing exons in addition to the backbone (Figure 2-5). The middle exon contains ~5 bases on each side that has weak splice (donor and acceptor) sites in addition to 30 bases containing the variant of interest in the middle, which need to exon splice enhancers for the splicing to take place. For each variant examined, there were WT and mutant plasmid that only differ in the middle nucleotide (SNV of interest).

The expressed product of WT and mutant plasmids were compared to check whether the reference and non-reference alleles produced different sized products, which would lead to the conclusion that the non-reference allele has an effect on an ESE. A drawn example of the expected results is shown in figure 2-6, which shows results of a complete loss or gain of splicing. However, two bands can also be observed instead of just single one which could indicate a reduction or increase in splicing efficiency rather than complete gain or loss of splicing.

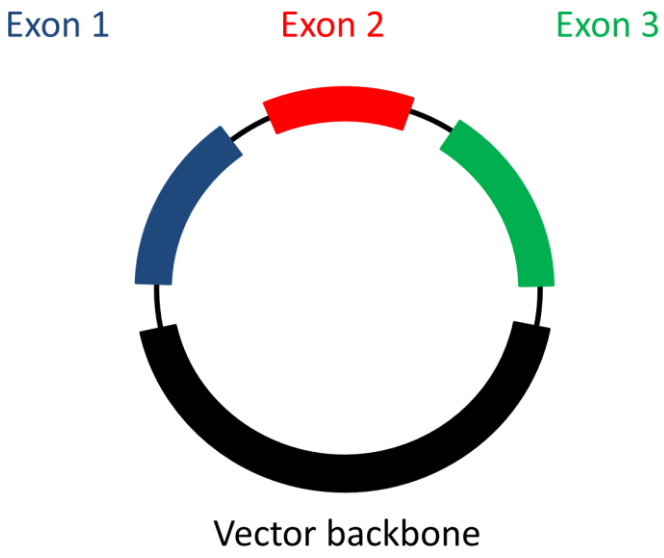


Figure 2-5: Diagram of the pcDNA-Dup (SF2-ASF3x) and PET01 plasmids.

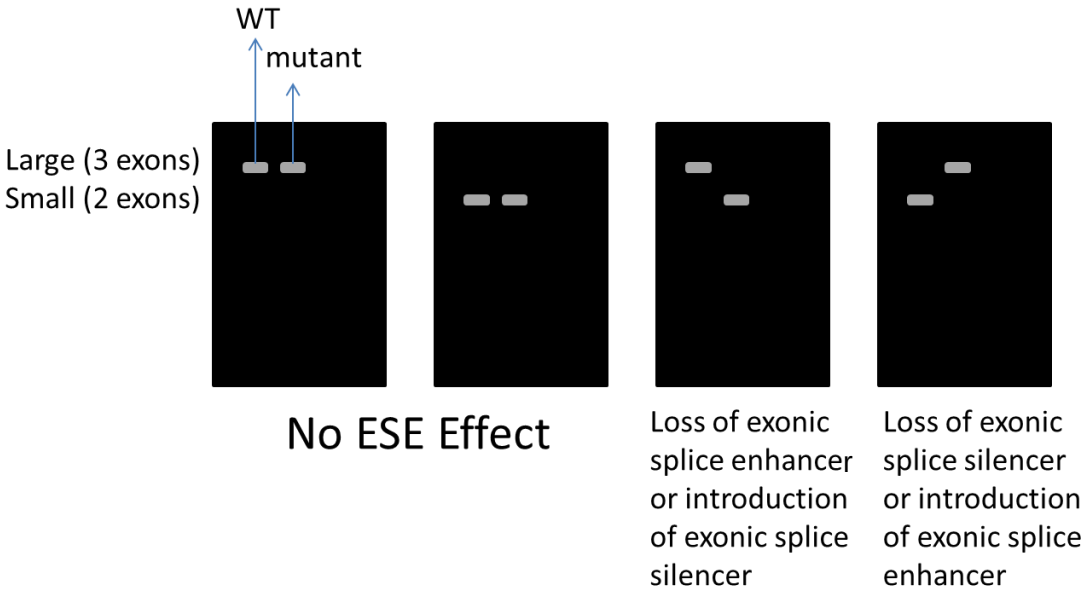


Figure 2-6: Sketch diagram illustrating expected results from the pcDNA-Dup (SF2-ASF3x) plasmid following transfection.

2.2.2.4.2 Acceptor / donor motif splicing assay

Acceptor / donor motif splicing assay uses a predesigned plasmid that aims to investigate the effect of a given SNV on splicing using PET01 plasmid. It uses a similar plasmid structure to pcDNA-Dup (SF2-ASF3x), however, the difference is that the insertion includes the whole exon and 300 bases of the flanking intron from each side to have better indication of the effect of SNV in reality.

In this project, it was used to assess the effect of *VWF* SNV on splicing. WT and mutant plasmids were prepared for each SNV. The expected outcome of this plasmid is shown in figure 2-7. Given that the whole exon is inserted, the outcome may vary depending on the position of the variant. For instance, there could be full or partial exon skipping, a full or partial intron included with the product or mix of these effects.

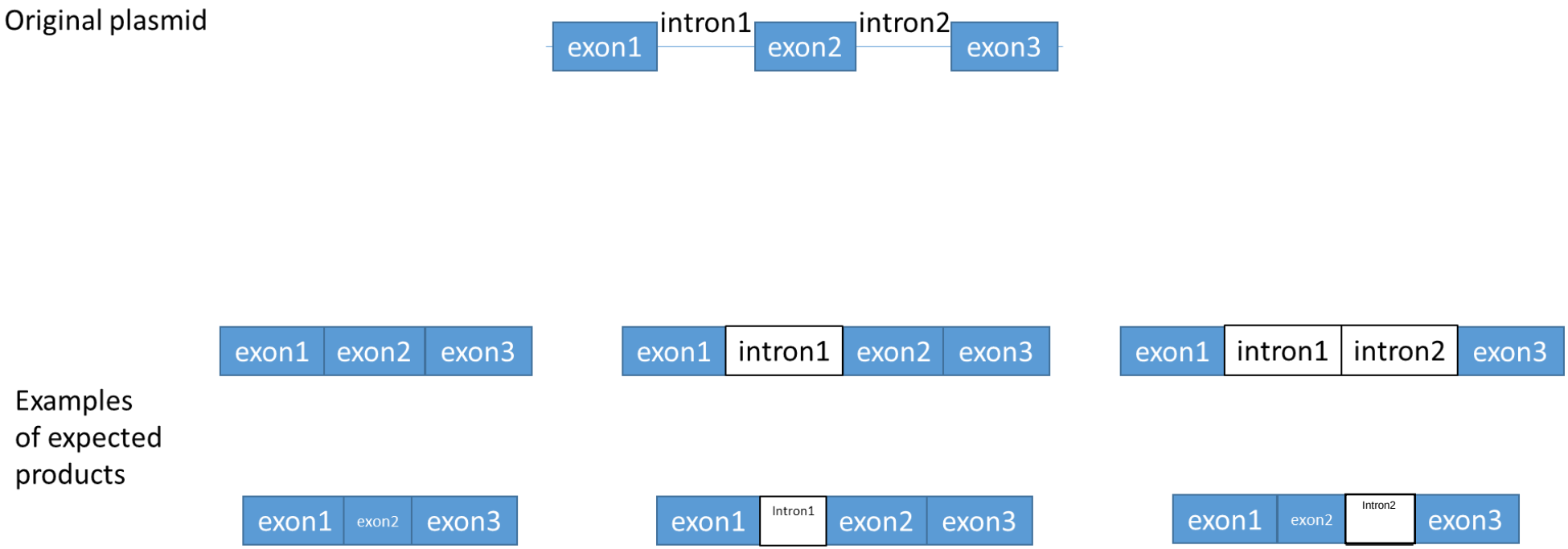


Figure 2-7: Expected results from the PET01 plasmid following transfection. Expected results could be products containing three exons, the middle exon being considered as an intron, an intron being considered as an exon or part of the intron or part of the exon only recognised as an exon.

2.2.2.4.2.1 Plasmid digestion

In order to allow ligation of the desired DNA sequence the pcDNA-Dup (SF2-ASF3x) plasmid was digested using *Eco*RI and *Bam*HI while the pET01 plasmid was digested using *Not*I and *Bam*HI (New Englands Biolabs) following the manufacturer's guidelines: 1µg plasmid DNA was digested using 1µl of each of the two enzymes and 5µl CutSmart® buffer in a 50µl reaction made up with dH_2O. The mixture was incubated at 37°C overnight and the enzymes were then deactivated at 65°C for 20 min. Following that, the digestion product was run on either 1% (PCR products) or 0.5% agarose gel (plasmids) (Section 2.2.1.7) in order to assess the digestion efficiency and specificity.

2.2.2.4.2.2 Insertion preparation

The insertion used for the pET01 plasmid was prepared by amplifying the region of interest in *VWF* using Q5 high fidelity polymerase (New England Biolabs). Gel extraction was then performed to eliminate any unwanted products (Section 2.2.1.9). Finally, the product was used for the ligation step (Section 2.2.2.2.2.3).

The insert for pcDNA-Dup (SF2-ASF3x) plasmid was a 30 base oligo (where the SNV of interest is in the middle). Complementary forward and reverse oligos were prepared by Eurofins and then were diluted to 20pmol. 5µl of each primer were mixed and annealed by being heated to 95°C in a heat block and left to cool down to 30°C.

2.2.2.4.2.3 Ligation

Ligation of the plasmid and insert was performed using T4 DNA ligase (Promega). Per ligation, 1µl T4 DNA ligase, 1.2µl ligation buffer, 10ng plasmid and 100ng PCR product/oligo were mixed and incubated at room temperature for 90 min. The enzymes were then deactivated at 70°C for 7 min. The ligated product was then transformed in NM554 competent cells (Section 2.2.2.1.6) followed by extracting the plasmid DNA using bacterial miniprep (Section 2.2.2.1.8). The samples were sequenced to check that the correct insertion had been obtained (Section 2.2.2.1.10). If the insertion was successfully

obtained, mutagenesis was performed (Section 2.2.2.1.5) to obtain the mutant plasmid.

2.2.3 Statistical analysis

The statistical analyses in this project were performed using GraphPad Prism v5.04 (GraphPad Software Inc., La Jolla, CA, USA). The t-test was used to compare any two groups (ordinary t-test to compare the mean, while the Mann-Whitney test was used to compare median). The One-way ANOVA was used to compare any three or more groups (ordinary One-way ANOVA used to compare the mean, while Kruskal Wallis test was used to compare the median) (Zhang and Zhang 2009). Tests performed on the population compared the median because VWF level is not normally distributed in the general population while tests performed on *in vitro* experiments compared the mean (Lethagen*, et al* 2008). A p-value <0.05 was considered significant.

3 Chapter three: Investigation of the association of polymorphic *VWF* SNV with VWF level

3.1 Background

Genetic variation within *VWF* has been associated with VWF plasma level in VWD patients as well as in normal individuals. Several genome wide association studies have identified variants within *VWF* and other genes to be significantly associated with VWF plasma level (Smith*, et al* 2010, Zabaneh*, et al* 2011). However, to date, there are very limited studies that have investigated the mechanism by which SNV influence VWF plasma level.

Sequence variants that have been reported by association studies in relation with VWF plasma level include c.2365A>G (p.Thr789Ala; rs1063856) and c.2385T>C (p.Tyr795=; rs1063857) both of which are located in exon 18 and are in perfect LD with each other (Lacquemant*, et al* 2000, Smith*, et al* 2010, van Schie*, et al* 2011a, Tang*, et al* 2015). A recent study assessed the association of rare and common VWF variants in African Americans, The study identified several variant with significant and independent effect on VWF level including p.Thr789Ala (Johnsen*, et al* 2013)

The aim of this chapter was to investigate *VWF* SNV and their association with VWF plasma level. Subsequently, the mechanism by which these SNV influence VWF level was investigated. It was hypothesised that SNV within *VWF* act as phenotypic modifiers; they may impact VWF level, but to a lesser extent than for a pathogenic mutation.

3.2 *In silico* analysis of *VWF* SNV

Genotyping of SNV across *VWF* and flanking regions had been conducted prior to the start of this project by Dr Christer Halldén (Lund University, Sweden). This genotyping study analysed 111 variants in ~1100 HC from the MCMDM-1VWD study using Sequenom genotyping (Appendix 1) (Johnsen*, et al* 2013). This data along with previously available phenotypic data including VWF:Ag level and ABO blood group (Goodeve*, et al* 2007) were analysed to identify any variants having a significant association with VWF plasma level. Eleven SNV (10%) were found to be significantly associated with VWF plasma level (Table 3-1). Four of these variants were located within *VWF* exons, four

were located within *VWF* introns, one was located between *VWF* and *ANO2* and two were located within *ANO2* (3' of the *VWF* locus).

Of the 11 SNV highlighted, c.2385T>C had previously been identified in the CHARGE association study to have a significant association with VWF level ($p=1.7x10^{-32}$) (Lacquemant*, et al* 2000, Smith*, et al* 2010). Another SNV highlighted during analysis of the genotyping data c.2365A>G was found to be in very strong LD with c.2385T>C (r=99.1%) and had also been reported in an earlier study showing a similar association with a ~12% increase in VWF plasma level (Zabaneh*, et al* 2011). These two variants were also shown to be in strong LD using the available HapMap data and were located 20bp apart in *VWF* exon 18 (Figure 3-1). Therefore, further investigation was carried out to investigate their association with VWF level.

An additional two SNV highlighted following analysis of the genotyping data rs11063953 (c.7888-2310T>C) and rs10849362 (c.7888-3501A>G) located 1.5kb apart in *VWF* intron 47 were both found to be significantly associated with ~15% and ~10% reduction in VWF plasma level respectively (with the non-reference allele) and were also investigated further for association with VWF level.

Although the remaining SNV showed a significant overall association with VWF level, they did not show consistent association with level when comparing the levels of reference allele, heterozygous and the non-reference allele (i.e. there was no gradual increase or decrease in VWF level between the three groups). Thus, these SNV were not further investigated (Table 3-1).

Table 3-1: *VWF* and *ANO2* SNV significantly associated with VWF plasma level in HC.

SNV	Nucleotide Change	a.a change	Location	Association with VWF level (median VWF:Ag; IU/dL)			p value[1]	LD (r^2)
				R/R	R/NR	NR/NR		
rs12580835	c.22+4578A>G		*ANO2* intron 1	91	99	95	0.010	99.7%
rs7310736	c.22+2097G>A		*ANO2* intron 1	91	99	95	0.007	
rs1063857	c.2385T>C	p.Tyr795=	*VWF* Exon 18	94	95	103	0.002	99.1%
rs1063856	c.2365A>G	p.Thr789Ala	*VWF* Exon 18	94	96	103	0.001	
rs11063951	c.-1342C>G		3' of *VWF*	98	94	83.5	0.035	99.5%
rs11063953	c.7888-2310T>C		*VWF* Intron 47	98	94	83	0.022	
rs216321	c.2555G>A	p.Arg852=	*VWF* Exon 20	97	89	95	0.005	92.8%
rs216339	c.2443-3627A>G		*VWF* Intron 18	97	88.5	95	0.007	
rs4764482	c.1533+2387A>G		*VWF* Intron 13	97	93	98	0.048	
rs10849362	c.7888-3501A>G		*VWF* Intron 47	98	95	90	0.022	
rs7312411	c.1548T>C	p.Tyr516=	*VWF* Exon 14	95	94	101.5	0.041	

*R= reference allele, NR= non-reference allele, LD= linkage disequilibrium, SNV= single nucleotide variant

[1] One-way ANOVA (Kruskal-Wallis test).

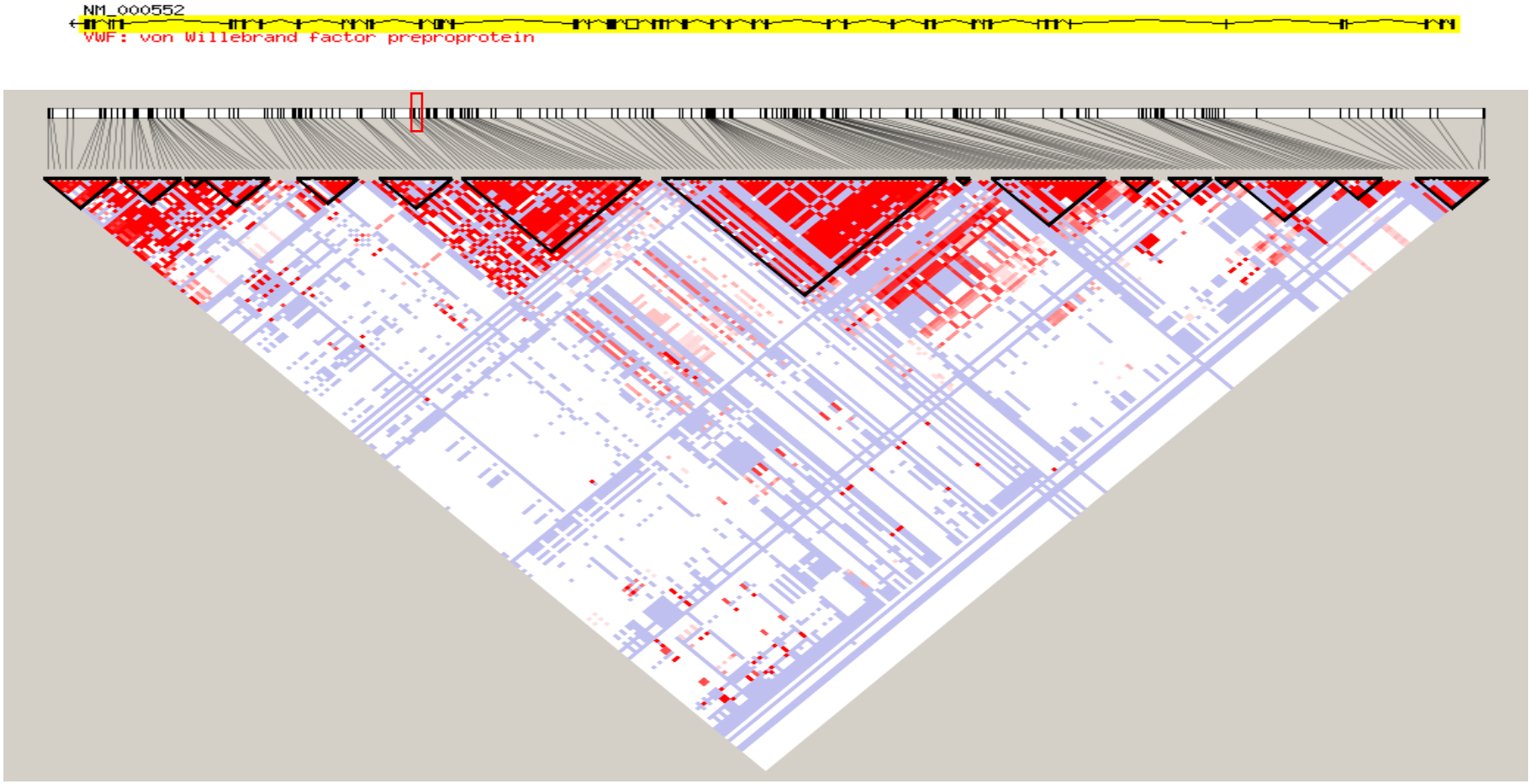


Figure 3-1: HapMap data of *VWF* gene. The LD between variants alleles is shown, in red square is the variant identified by the CHARGE study. White squares indicate weak LD, red squares indicate strong LD while Grey boxes indicate a lack of data for any given SNV association.

3.2.1 *In silico* analysis of c.7888-3501A>G and c.7888-2310T>C

Although c.7888-2310T>C and c.7888-3501A>G were significantly associated with VWF plasma level, they were not in strong LD with one another (r^2= 34%) which suggested that they may have an independent effect on VWF plasma level. These two variants were analysed using various prediction tools to try and predict their mechanism of action. SNV within 3kb of each of the two variants were investigated as well to check whether the observed association was because of these two specific SNV or other surrounding SNV that were potentially in strong LD with one or both variants.

According to the prediction tools, c.7888-3501A>G seemed to be introducing a new donor splice site (by 4 out of 5 tested tools) within intron 47, therefore the observed association was potentially due to this variant. However, c.7888-2310T>C was not predicted to have any major effect on splicing, ESE motifs, miRNA binding sites or transcription factor binding sites (Table 3-2). Therefore, the observed association could be due to another variant in strong LD with this variant. However, no other variant was found to be in strong LD with c.7888-2310T>C using the available genotyping data set, which was missing a large proportion of known *VWF* variants due to it being selective and covering only a small proportion of SNV throughout *VWF*. Therefore, HapMap data was used to test the LD in that region further, however, this variant still did not show strong LD with any other variant in that region. This in turn suggested that c.7888-2310T>C is either in strong LD with a variant(s) not covered in the HapMap data or it has an effect that could not be predicted using the *in silico* tools. Due to time limitations, c.7888-3501A>G and c.7888-2310T>C were excluded from further investigation.

Table 3-2: *In silico* analysis of variants surrounding and including c.7888-3501A>G and c.7888-2310T>C.

	ESE effect		Splicing effect					Transcription factors				
Variant	**HSF ESE**	**ESE finder****	**HSF splice**	**fruitfly**	**Netgen2**	**ASSEDA***	**ASSP***	**Promo****	**MET**	**miRNA****	**Rs number**	**MAF**
c.7888-17T>C	Minor	1	No effect	Loss acceptor	No effect			1	Minor	0	rs12297370	0.02
c.7888-93G>A	Major	0	Minor	No effect	No effect			0	No effect	3	rs3759320	0.4181
c.7888-118C>T	Minor	0	No effect	No effect	No effect			-3	Minor	1	rs3759321	0.4391
c.7888-129T>C	Major	2	No effect	Minor	No effect			0	Minor	0	rs12297442	0.2234
c.7888-382G>A	Minor	1	Minor	No effect	No effect	minor		0	Minor	0	rs12317523	0.26
c.7888-405G>T	Major	1	Major	New acceptor	No effect			0	Major	0	rs7969672	0.2398
c.7888-669T>G	Major	1	Major	No effect	New acceptor			0	Minor	-1	rs10774387	0.2424
c.7888-1016_7888-1015insT	Minor	0	Major	Minor	No effect	NA		0	Major	0	rs143246250	0.0339
c.7888-1145C>T	Major	2	Minor	Minor	No effect	minor		0	Major	1	rs723188	0.2762
c.7888-1199G>A	Major	2	Major	No effect	No effect	No effect		1	Minor	0	rs723189	0.3808
c.7888-1214A>G	Major	0	Major	No effect	No effect	No effect		0	Minor	-1	rs723190	0.0785
c.7888-134T>C	Major	0	Major	No effect	No effect			0	Major	1	rs57040304	0.0984
c.7888-1421_7888-1420insTG	Major	1	Major	No effect	No effect			0	No effect	0	rs145581063	0.0337
c.7888-1514_7888-1513insT	Major	0	Major	No effect	No effect			0	Minor	0	rs140939660	0.0781
c.7888-1603C>T	Major	1	No effect	No effect	No effect			0	Minor	0	rs114446920	0.0785
c.7888-1626A>G	Major	0	No effect	No effect	No effect			0	Minor	0	rs10849360	0.4111
c.7888-1654C>T	Minor	0	Minor	No effect	No effect			0	Minor	-1	rs78329863	0.1667
c.7888-1686C>T	Major	2	No effect	No effect	No effect			0	Major	0	rs12305676	0.2821
c.7888-1789G>T	Major	1	Major	No effect	No effect			0	Minor	-3	rs12319598	0.2821
c.7888-184G>C	Major	0	Major	No effect	No effect			0	No effect	0	rs115804723	0.0218
c.7888-1932C>T	Minor	1	No effect	No effect	No effect			0	No effect	0	rs12819922	0.2173
c.7888-223A>T	Major	1	Major	New acceptor	No effect			-2	Minor	0	rs7958883	0.381

	ESE effect		Splicing effect					Transcription factors				
Variant	**HSF ESE**	**ESE finder****	**HSF splice**	**fruitfly**	**Netgen2**	**ASSEDA***	**ASSP***	**Promo****	**MET**	**miRNA****	**Rs number**	**MAF**
c.7888-2310T>C	Major	1	No effect	No effect	New donor	Minor		0	Major	0	rs11063953	0.22
c.7888-2564C>T	Major	1	No effect	No effect	No effect	Minor		0	Minor	0	rs10774388	0.2033
c.7888-2856C>T	Major	2	No effect	No effect	No effect	Minor		0	Minor	0	rs117889430	0.013
c.7888-2991A>G	minor	-1	No effect	No effect	No effect			-1	Major	0	rs10849361	0.2039
c.7888-3047delC	Major	0	Major	Stronger new acceptor	No effect			1	Minor	-1	rs199571870	NA
c.7888-3124T>G	Major	0	No effect	No effect	No effect			-1	Minor	-1	rs115726920	0.0196
c.7888-3274A>C	Minor	1	No effect	No effect	No effect			0	Minor	0	rs76879130	0.0146
c.7888-3372T>C	Minor	1	No effect	No effect	No effect			0	Minor	-1	rs78818158	0.0793
c.7888-3501A>G	Minor	1	Major New donor	Strong New donor	No effect	New donor	New donor	0	Minor	0	rs10849362	0.4213
c.7888-3597T>C	Minor	1	No effect	No effect	No effect			0	No	0	rs10849363	0.4996
c.7888-3779T>C	Major	2	No effect	No effect	No effect			-2	Major	7	rs78245005	0.0793

*Websites that give limited access, empty boxes refer to unanalysed variants. ** + refers to the introduction of a new site while – refers to the loss of a previous site.

3.2.2 *In silico* analysis of c.2385T>C and c.2365A>G

Given the strong LD and close proximity of c.2385T>C and c.2365A>G, the observed association with VWF level could theoretically be due to just one of the two variants or a combined effect of the two variants together.

Before proceeding with *in vitro* experiments, the effect of the two variants was tested using various prediction tools. They were tested for any effect on protein or mRNA structure, function or splicing. Apart from SNV c.2385T>C showing a minor effect on ESE motifs and SNV c.2365A>G predicted to be damaging to VWF (using Align GVGD; Table 3-3), neither of these two variants were predicted to have any major effect.

Codon usage, which compares the abundance of different codons that encode the same a.a within the human genome or a given gene, was also investigated for the silent variant c.2385T>C. This showed that the non-reference codon was more abundant compared to the reference codon within the human genome as well as within *VWF* (Table 3-4) suggesting that the more non-reference allele could result in a faster translation rate, due to its abundance, and thus a higher VWF level.

Table 3-3: Predicted effect of c.2385T>C and c.2365A>G on the VWF protein and VWF RNA.

SNV	VWF protein			VWF RNA			
	Align GVGD	SIFT	PolyPhen	HSF Splicing	NetGen2	Fruitfly	ESE Finder
c.2365A>G	Most likely damaging	Benign	Benign	Benign	Benign	Benign	No effect
c.2385T>C	NA*	NA	NA	Benign	Benign	Benign	Minor effect

* NA: not analysed; prediction tools were not able to analyse synonymous changes.

Table 3-4: Tyrosine codon usage statistics.

Genome	Codon	Codon /1000
Human genome	TAT	12.2
	TAC	15.3
VWF	TAT	9.8
	TAC	17.3

* The values represent the number of codons present per 1000 codons.

3.3 Association of c.2385T>C and c.2365A>G with VWF level in the HC population

The two SNV were previously associated with VWF level, further analysis of association was carried out specifically in a European healthy control (HC) population to further confirm the observed association. The association of the two variants with VWF plasma level was analysed in 1035 HC from the MCMDM-1VWD study. Individuals with the non-reference alleles for both SNV were found to have significantly higher levels of plasma VWF by ~9% when compared to individuals with the reference allele (Figure 3-2).

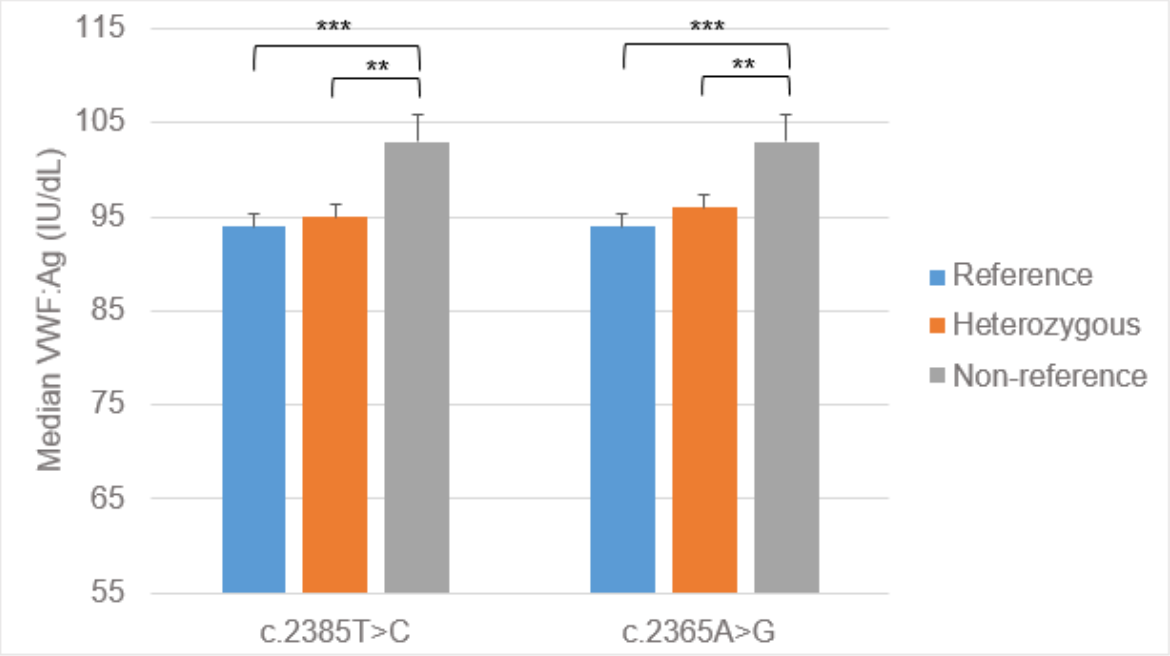


Figure 3-2: Association of c.2385T>C and c.2365A>G with VWF plasma level in 1035 HC. Significant differences were observed for c.2385T>C (p=0.003) and c.2365A>G (p=0.002) using the One-way ANOVA (Kruskal-Wallis test). The t-test (Mann-Whitney test) was used to compare genotypes (**≤0.005, ***≤0.0005). Bars indicate standard error.

3.4 Association of c.2385T>C and c.2365A>G with FVIII:C level in the HC population

These two SNV also showed significant association with FVIII:C level similar to that observed with VWF plasma level when examined in the 1030 HC. These two SNV are significantly associated with ~9% higher level of FVIII:C in the normal population (Figure 3-3).

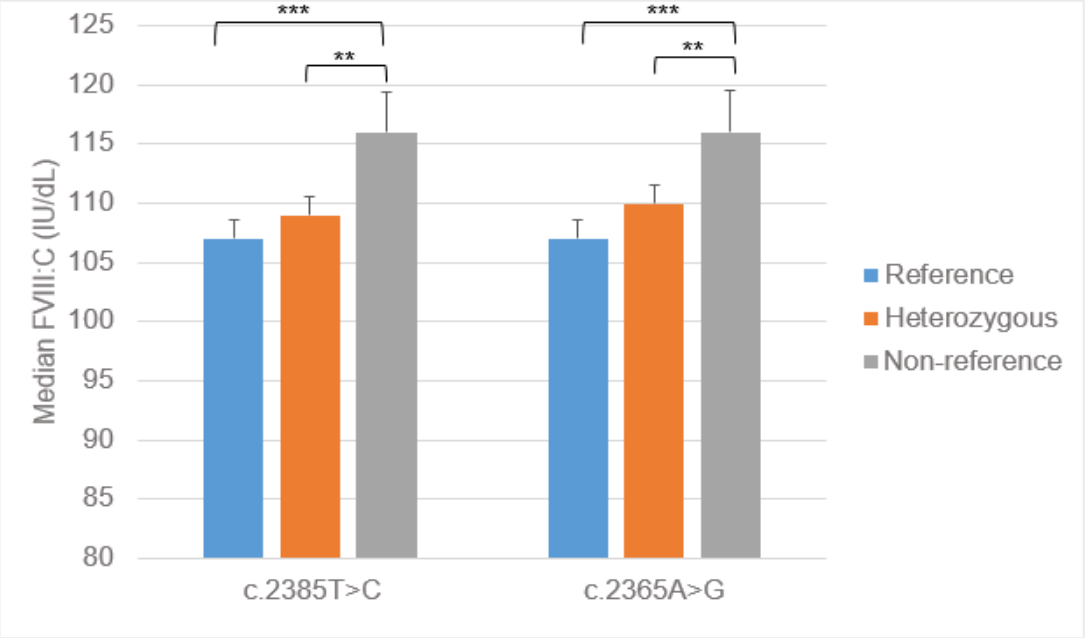


Figure 3-3: Association of c.2385T>C and c.2365A>G with FVIII:C plasma level in 1030 HC. Significant differences were observed for c.2385T>C (p=0.007) and c.2365A>G (p=0.005) using the One-way ANOVA (Kruskal-Wallis test). The t-test (Mann-Whitney test) was used to compare genotypes (**≤0.005, ***≤0.0005). Bars indicate standard error.

3.5 *In vitro* study of the effect of c.2385T>C and c.2365A>G

In order to investigate the mechanism by which c.2385T>C and c.2365A>G were influencing VWF level, *in vitro* studies were conducted on expressed VWF mRNA and protein. The independent effect of these SNV on VWF expression was investigated as well as their combined effect to determine whether one or both SNV were associated with VWF level.

3.5.1 Protein quantification

1.1.1.1 Transfection optimisation

The initial optimisation of transfection was performed within this research group previously (Ashley Cartwright, previous PhD student). However, some of the generated results showed wide variation. This led to the investigation of possible ways of improving the transfection protocol.

After experimenting with the protocol, no differences in renilla reading (used to normalise for transfection efficiency) were observed between empty wells and wells containing samples (Figure 3-4) suggesting that the sample plate and/or the renilla assay kit used were not appropriate. After reviewing the literature, neither were found to be appropriate. Previously used full white plates were replaced with white plates with a clear bottom to allow the plate reader to measure the samples. In addition, the assay kit used (Dual-Luciferase kit) was not the right option as it required measurement of firefly luciferase before renilla could be measured. It was replaced with the Renilla Luciferase Assay System which measures renilla luciferase only.

Value	1	2	3	4	5	6	7	8	9
A	4.34412	4.32365	4.28123	4.26576	4.29893	4.27546	4.27006	4.28162	4.31451
B	4.31646	4.26524	4.29587	4.29054	4.25335	4.24175	4.25361	4.2423	4.24265
C	4.28482	4.27912	4.25544	4.24246	4.24382	4.24702	4.2465	4.21383	4.28824

Figure 3-4: Renilla readings. Random samples in row A and B and empty wells in row C. There were no differences between the values of empty well or wells with samples.

1.1.1.2 Protein quantification of c.2385T>C and c.2365A>G

VWF protein level was measured after transfection of 4 different VWF expression plasmids (incorporated into pcDNA3.1/Hygro (-)) into a HEK293T cell line; wild-type (WT) plasmid, a plasmid containing the c.2385T>C change, a plasmid containing c.2365A>G and a plasmid containing both variants. Co-transfection of renilla was performed which was used for normalisation (section 2.2.2.3.1 for transfection and section 2.2.2.3.2 for protein quantification).

Quantification of VWF level was carried out to measure VWF protein in HEK293T cells transfected with WT VWF plasmid or VWF plasmid carrying the desired change (i.e. c.2385T>C and c.2365A>G). The results indicated that the non-reference allele of each SNV was associated with a significantly increased level of expression of VWF both within the cell and secreted from the cell. The plasmid containing both variants showed an additive increase in VWF protein expression (Figure 3-5).

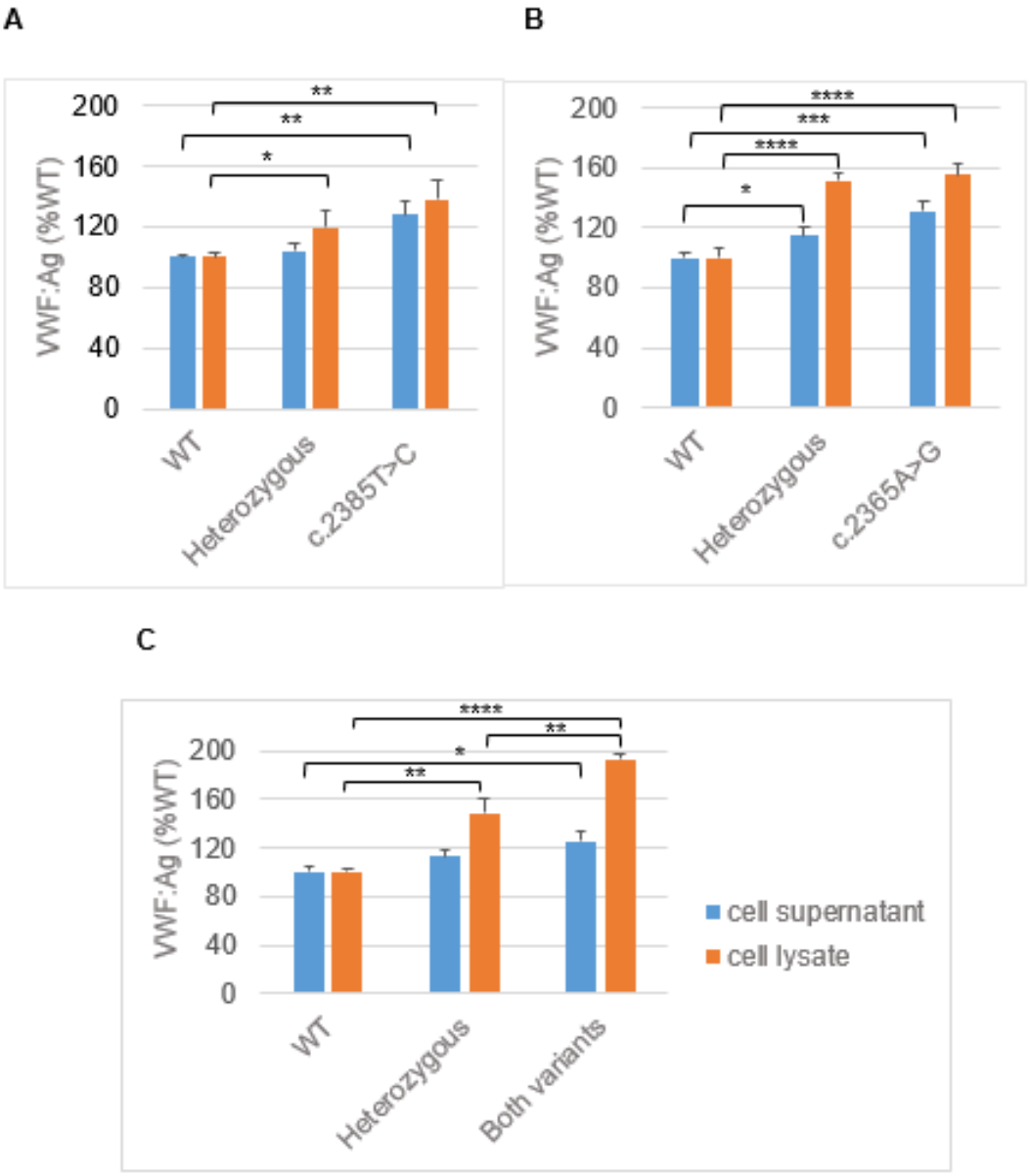


Figure 3-5: VWF expression in cell lysate and supernatant for c.2385T>C, c.2365A>G and both variants *in cis* compared to WT and heterozygous plasmids. Significant differences were observed for **A)** c.2385T>C: cell supernatant (p=0.006) and cell lysate (p=0.03), **B)** c.2365A>G: cell supernatant (p=0.0009) and cell lysate (p<0.0001), **C)** both variants: cell supernatant (p=0.03) and cell lysate (p<0.0001) using the ordinary One-way ANOVA. The t-test (Mann-Whitney test) was used to compare genotypes (*≤0.05, **≤0.005, ***≤0.0005 and ****≤0.00005). The experiment was repeated three independent times. Bars indicate standard error.

3.5.2 RNA quantification

Given that the non-reference allele of each of c.2385T>C and c.2365A>G led to increased protein levels, the effect of both SNV on mRNA expression was investigated. RNA quantification was carried out on plasmids containing each SNV independently and together to study their independent and combined effect on VWF RNA expression level.

1.1.1.3 qPCR optimisation

Optimisation of qPCR went through various steps to ensure that reliable and reproducible data was generated as follows:

Singlex master mix was initially used following the standard protocol (Section 2.2.2.1.123). *VWF* and *B2M* (endogenous control) RNA were measured for each sample in separate wells. However, the data generated showed a large variation which was confirmed by large variation in standard error (Figure 3-6).

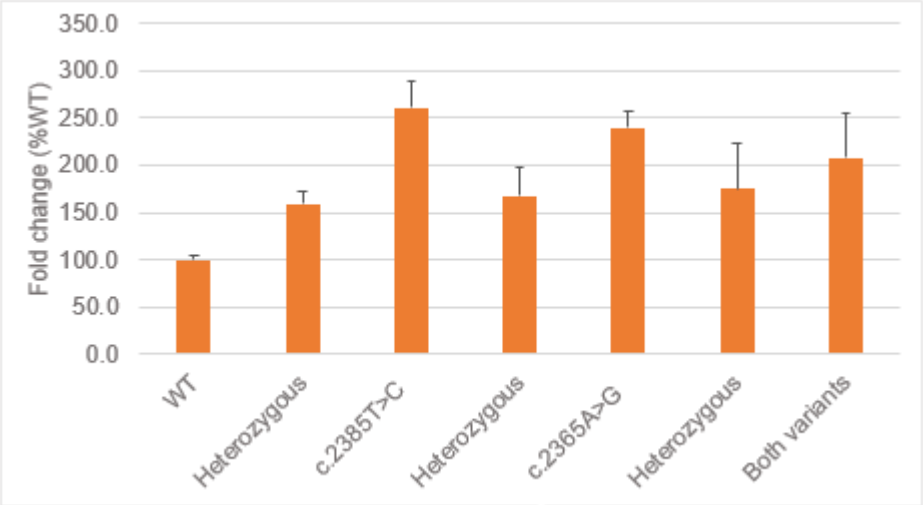


Figure 3-6: VWF RNA expression using plasmid expressing c.2385T>C, c.2365A>G and both variants compared to WT and heterozygous expression. WT: 100%; heterozygous (c.2385T>C): 159.4%; c.2385T>C 261%; heterozygous (c.2365A>G): 167.4%; c.2365A>G: 239.7%; heterozygous (c.2385T>C and c.2365A>G): 175.2%; both variants: 207.9%. The experiment was repeated three independent times. The bars represent standard error.

The experiment was repeated using the same protocol but measuring *VWF* and *B2M* in the same well to reduce this variation. However, as *VWF* was over expressed (compared to the endogenous control) due to the expression plasmid, the nucleotides were exhausted before *B2M* was detected. This was shown by the *B2M* curve not reaching its peak after either 40 or 50 cycles (Figure 3-7). Therefore the suggestion (as recommended by the manufacturer) was to use the multiplex mastermix which has concentrated nucleotides.

Measuring the RNA of *VWF* and *B2M* in the same well was repeated using a multiplex mastermix. However, the data generated could not be analysed. The multiplex mastermix was found to use a different reference dye (mustang purple), which is used to normalise the detection against variables including air bubbles or evaporation, compared to the dye ROX that is used in singlex mastermix. Due to the available qPCR machine not being calibrated for the mustang purple reference dye, the alternate option was to try using singlex mastermix without adding any dH_2O to the mix and instead add more mastermix. This would provide more nucleotides for the reaction. In addition, primer-limiting probes for *VWF* were also used, which provides a limited quantity of VWF primers in order to limit its amplification giving more chance to amplify and detect *B2M*.

The experiment was repeated using these conditions and the results suggest that both *VWF* and *B2M* were amplified and detected before the end of the 40^{th} cycle (Figure 3-7). The significant improvement in amplification reproducibility demonstrated that the qPCR was now optimised.

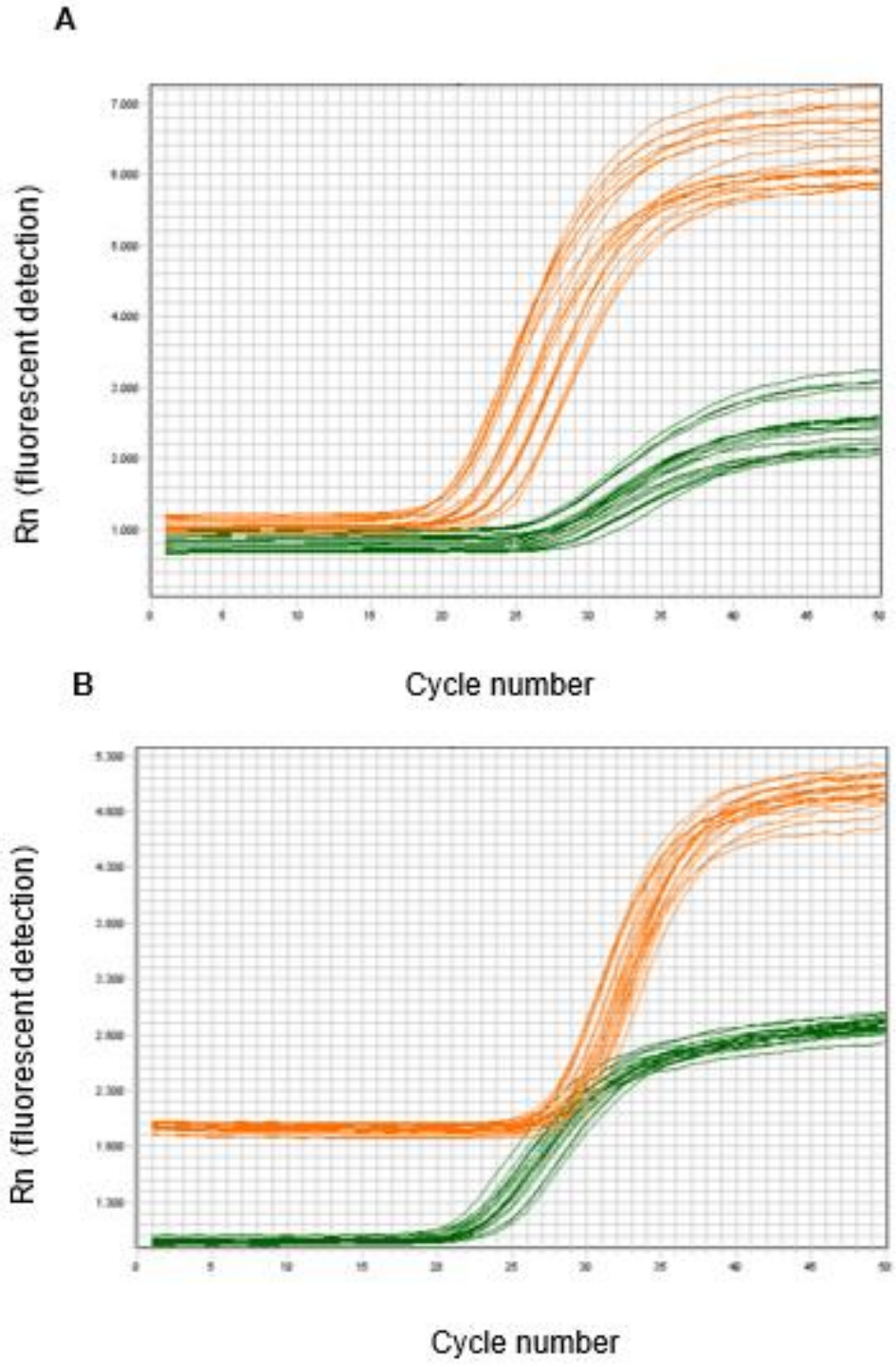


Figure 3-7: Detection of *VWF* RNA expression and *B2M* during 50 PCR cycles. A) Within the first 40 cycles, *VWF* (orange) peaked (exhausting the nucleotides), while even after 50 cycles, *B2M* (green) did not peak. **B)** Optimised conditions allowed for early peak and detection (before the 40th cycle) of both *VWF* (green) and *B2M* (orange).

1.1.1.4 RNA quantification of c.2385T>C and c.2365A>G

As with protein expression, quantification of c.2385T>C and c.2365A>G RNA was carried out to test their independent and combined effect on RNA expression. The results supported the previous findings for protein expression. Furthermore, both variants were again shown to have an additive effect on RNA expression (Figure 3-8).

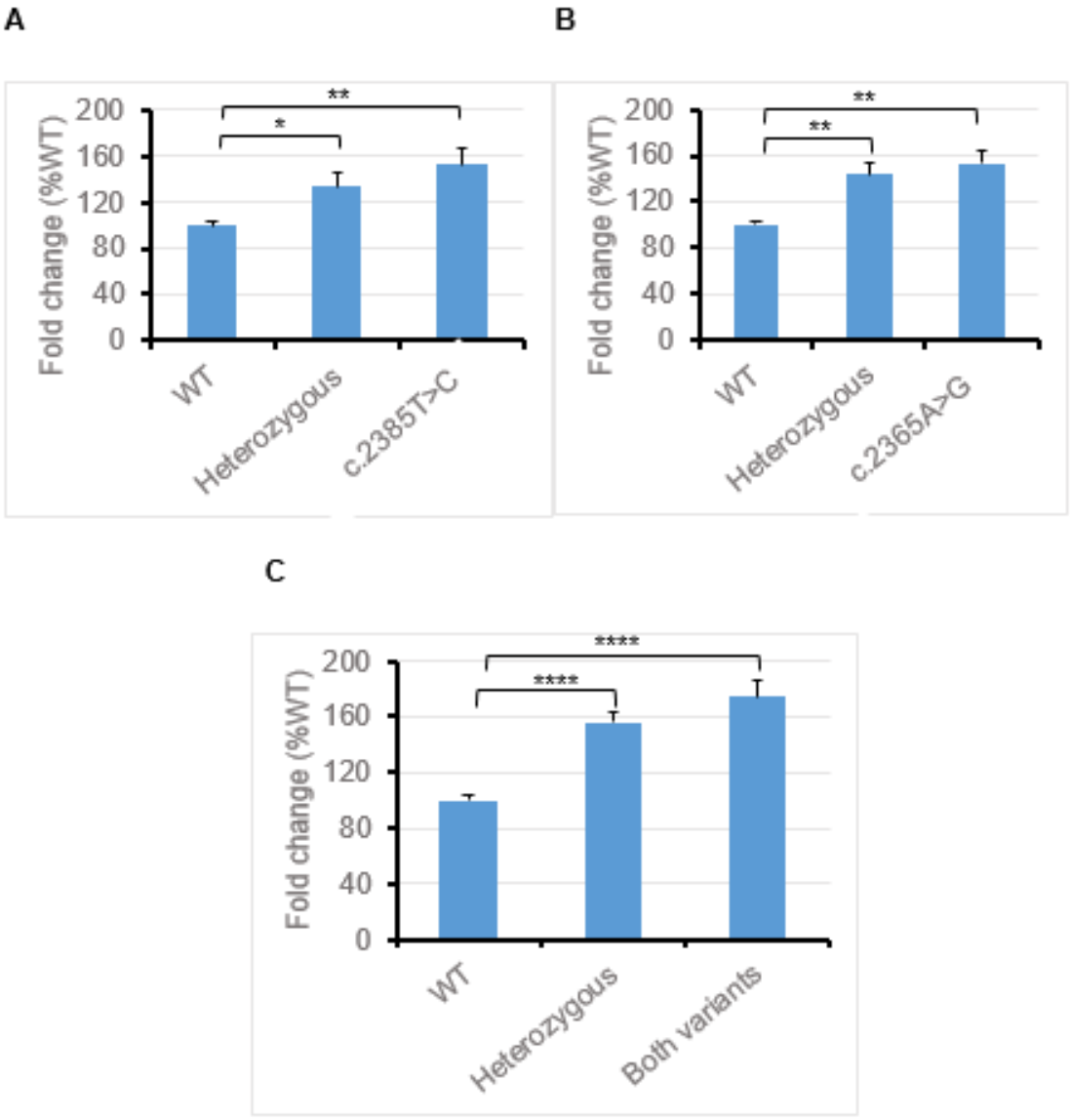


Figure 3-8: RNA expression of c.2385T>C, c.2365A>G and both SNV compared to WT and the heterozygous variant. Significant differences were observed for **A)** c.2385T>C (p=0.02), **B)** c.2365A>G (p=0.005), **C)** both variants (p<0.0001) using the ordinary One-way ANOVA. The t-test (Mann-Whitney test) was used to compare genotypes (*≤0.05, **≤0.005 and ****≤0.00005). The experiment was repeated three independent times. Bars indicate standard error.

3.5.3 mRNA splicing analysis

Variant alleles of c.2385T>C and c.2365A>G showed significantly higher VWF protein and mRNA expression. One of the possible mechanisms that might cause this was that the reference allele for both variants could be causing a slight defect to the splicing mechanism thus reducing the VWF level compared to the non-reference allele. Even though neither variant had been predicted to have any effect on splicing using various *in silico* tools, the experiment was carried out as these tools are only used for an indication and they are not error proof. To test for a possible effect on splicing, an ESE assay and an acceptor / donor motif splicing assay were used to investigate the effect of the two variants on ESE motifs and mRNA splicing respectively.

3.5.3.1 Exon splice enhancer (ESE) assay

In this project, a pcDNA-Dup (SF2-ASF3x) plasmid was used to assess the effect of c.2385T>C and c.2365A>G on ESE (Section 2.2.2.2.1).

3.5.3.1.1 Optimisation

The oligos for the minigene were designed to include the 30 bases of interest (the variant of interest and 15 bases each side), digestion sites and M13 tails to give enough space for the digestion enzyme to bind and perform the digestion. The standard protocol (Section 2.2.2.2.2.1) was followed with digestion of the oligos and the plasmid in addition to ligation using T4 DNA ligase. The final product was transformed into *E.coli* NM554 competent cells. However, no colonies were found even after changing the digestion time of the plasmid from 1.5h to 4h. New enzymes were ordered and the experiment was repeated.

Using the new enzymes and the same conditions, colonies were obtained, DNA was extracted and sent for sequencing. The sequence revealed that the majority of the colonies carried the WT plasmid without any modifications. However, some plasmids showed digestion ~1.1kb downstream of the expected site. Therefore, new primers were designed to sequence this area to check for the presence of another digestion site for the enzymes that were used. Sequencing that area in the original plasmid revealed a site with a 4/5 nucleotide match to the digestion site. Further investigation showed that the reaction buffer used was not 100% compatible with both restriction enzymes which could explain the non-specificity in the observed digestion in the sequence obtained as well as an extra band when it was run on a gel (Figure 3-9). Therefore, a reaction buffer 100% compatible with both enzymes was ordered and the experiment was repeated.

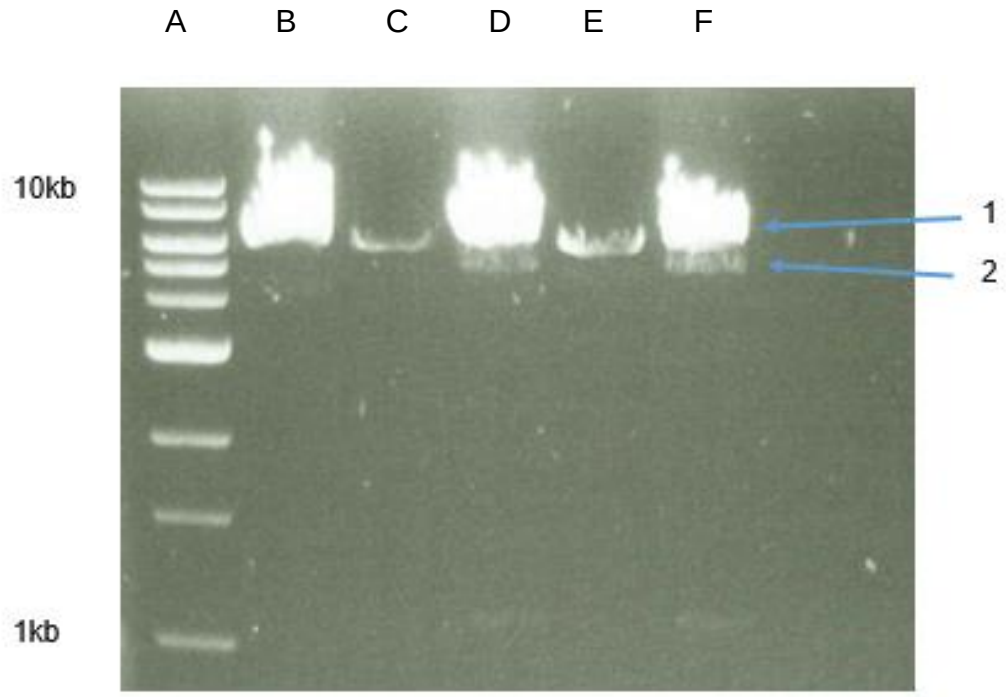


Figure 3-9: Digestion of pcDNA-Dup (SF2-ASF3x) plasmid. Digestion performed using new enzymes and old reaction buffer showing an extra band (2) in D and F below the main band (1). A) ladder; B) single digest; C) 0.5µl of double digest for 30 min; D) 2µl of double digest for 30 min; E) 0.5µl of double digest for 2h; F) 2µl of double digest for 2 h.

Colonies were obtained and the DNA was sent for sequencing. However, the sequence returned was either WT or empty plasmid (i.e. plasmid without the middle exon) suggesting that the digestion succeeded but the ligation failed possibly due to insufficient oligo being present. Therefore, double the amount of oligo was added to the ligation step. However, the results remained the same suggesting that the oligo might not be digested for long enough. Therefore, digesting the plasmid and oligos overnight was performed.

The oligo and plasmid were digested overnight and run on a polyacrylamide gel and on an agarose gel respectively. The aim was to test the digestion efficiency between 4h and overnight. The results suggested that overnight digestion resulted in much better efficiency with more double digestion observed in the oligo (Figure 3-10). These newly digested oligos were then used with the same protocol to perform the ligation. However, the results remained the same. The obtained colonies either had WT plasmid or empty plasmid. Further modification to the protocol included gel extraction of the digested plasmid before proceeding to the ligation step in order to limit the presence of WT (undigested) plasmid from the digested WT and limit their re-ligation during the ligation process.

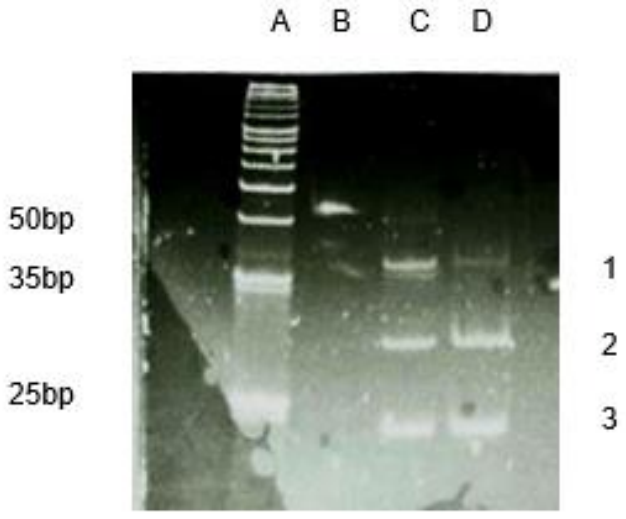


Figure 3-10: Comparison between digestion of oligos for 4 h and overnight. A) ladder; B) undigested oligo; C) 4hr digested oligo; D) overnight digested oligo. 1) single digestion (~40bp); 2) double digestion (~30bp); 3) digested product (~20bp)., electrophoresed on a 4% polyacrylamide gel.

After repeating the experiment using the newly digested plasmid and oligo, and gel extracting the digested plasmid, the sequence of the obtained colonies showed an empty plasmid but no WT plasmid suggesting that the gel extraction step is important. However, it did not completely solve the problem.

During the previous optimisation steps, items improved included using new enzymes with a 100% compatible reaction buffer, a greater quantity of oligo to increase the chance of ligation, better digestion overnight and gel extracting the plasmid to eliminate the possibility of obtaining WT plasmid. However, no successful cloning was obtained. This would lead to two possibilities (Figure 3-11) which are; 1) the oligo is not ligating at the desired site, 2) the oligo is ligated but then the plasmid is ligating again with itself. Therefore, new oligos were ordered that did not require digestion (i.e. they had the 30 bp of interest and the required digestion overhangs) and the whole experiment was repeated.

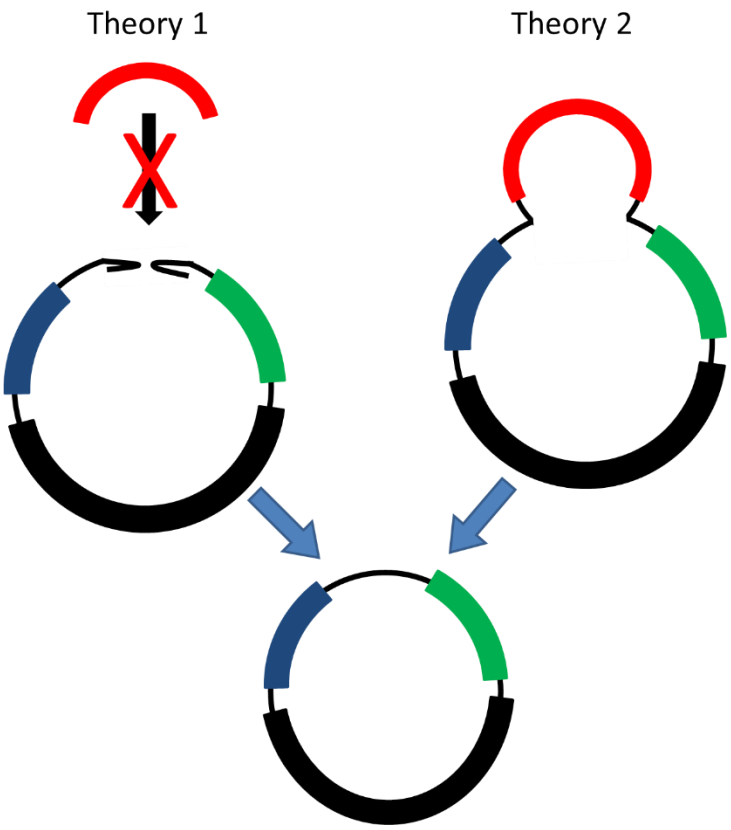


Figure 3-11: Possible reason for the non-successful cloning. Theory one shows that the oligo did not bind to the plasmid so that the plasmid self-ligated. In theory two, the oligo was successfully ligated, then the plasmid re-ligated with itself leading to the same observed result.

However, the obtained results remained the same. Thus, assuming the 2nd theory was right, ligation at 3 h was too long. Therefore, the experiment was repeated and the ligation time was reduced to 15 or 30min.

The experiment was repeated using the original protocol with the following changes, overnight digestion of the plasmid, pre-digested oligo, double the quantity of the oligo, gel extracting the digested plasmid, and reducing the ligation time to 15min and 30min. Colonies were obtained, DNA was extracted and sent for sequencing. Finally, successful insertion of the middle exon was obtained (Figure 3-12).

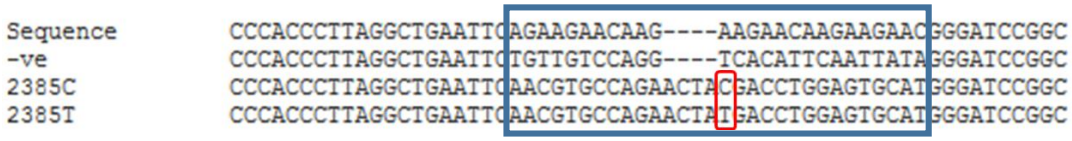


Figure 3-12: Sequence of the middle exon for original plasmid pcDNA-Dup (showing the +ve control (referred to as sequence), -ve control and both alleles for c.2385T>C). Outlined in red is the nucleotide difference between reference and non-reference allele. Outlined in blue is the middle exon insertion.

3.5.3.1.2 Effect of c.2385T>C and c.2365A>G on ESE

Following the previous optimisation, the experiment was performed on c.2385T>C and c.2365A>G using the obtained plasmids (including WT and mutant for each variant separately and in the same plasmid as well as positive and negative controls). The results suggest that these variants do not affect ESE when comparing the band of the reference and non-reference alleles which were of identical size (Figure 3-13).

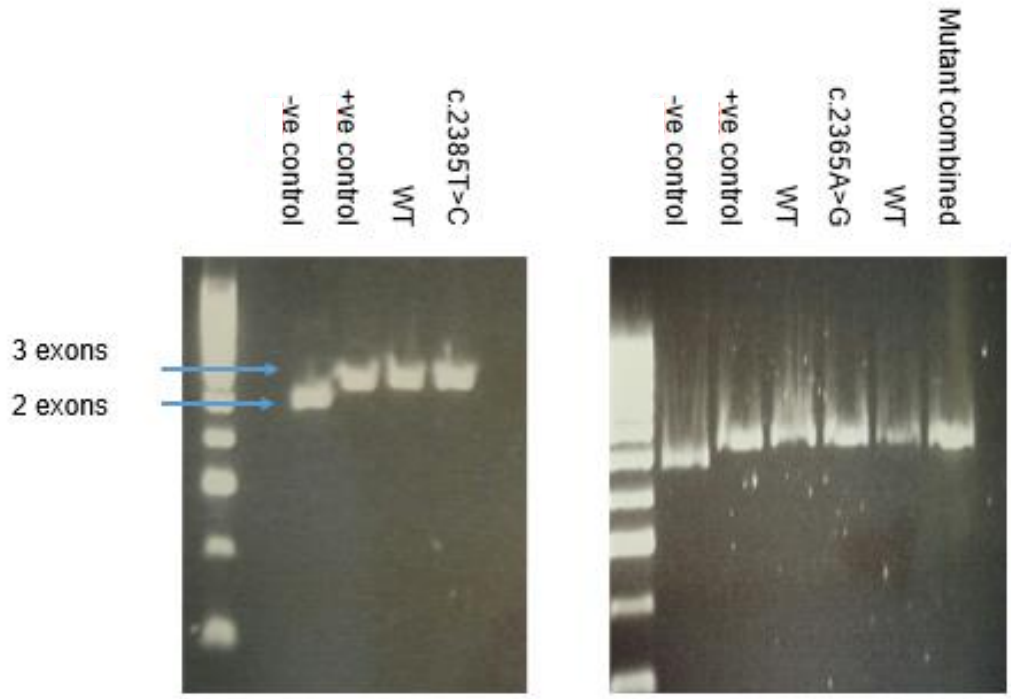


Figure 3-13: ESE assay results for c.2385T>C and c.2365A>G. Similar band sizes were observed in WT and mutant plasmids suggesting no effect on ESE. -ve control, plasmid with an insertion that is known not to express the middle exon (thus showing a small product size with only 2 exons); +ve control plasmid with an insertion that is known to be expressed (thus showing a larger product size 3 exons).

3.5.3.2 Acceptor / donor motif splicing assay

In this project, pET01 plasmid was used to assess the effect of c.2385T>C and c.2365A>G on mRNA splicing. Therefore, four plasmids were made in addition to the +ve and –ve controls (Section 2.2.2.2.2). These plasmids included a WT exon 18 insertion, exon 18 insertion with c.2385T>C, exon 18 insertion with c.2365A>G and exon 18 insertion with both c.2385T>C and c.2365A>G. Following preparation, these plasmids were transfected into HEK293T cells, mRNA was extracted and cDNA was prepared and amplified using specific primers. The results were then analysed in the same way as for the ESE assay where the product size was compared between WT and mutant plasmids. The results showed no differences between WT and any of the three mutant plasmids suggesting no effect of these variants on mRNA splicing (Figure 3-14).

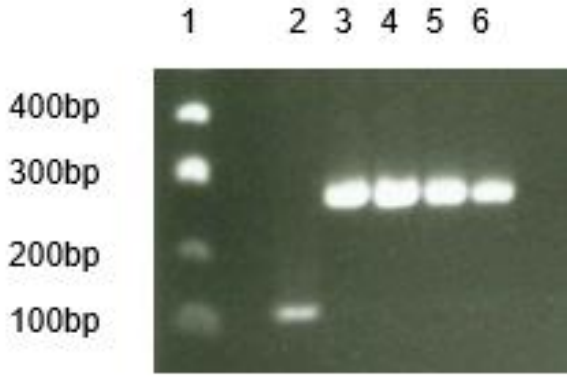


Figure 3-14: The effect of c.2385T>C and c.2365A>G on splicing. 1: ladder; 2: -ve control missing the middle exon; 3: exon 18 WT; 4: exon 18 with c.2385T>C; 5: exon 18 with c.2365A>G; 6: exon 18 with both c.2385T>C and c.2365A>G. Exon 18 size is ~160bp, -ve control size ~100bp, final product size ~260bp.

3.6 mRNA half-life

The effect of c.2385T>C and c.2365A>G on mRNA half-life was then investigated. Previously prepared VWF plasmid (for protein and mRNA quantification; section 3.5.1) was used for this experiment. The aim was to compare mRNA half-life between WT plasmid and plasmid containing either or both of the variants. mRNA half-life refers to the amount of time needed for 50% of the mRNA to be degraded. Actinomycin D treatment of the cells was used for this purpose as it inhibits the production of mRNA (Section 2.2.2.1.12.4).

WT plasmid was transfected into HEK293T cells then treated with 2.5µg/ml, 5µg/ml and 10µg/ml of actinomycin D 24h post transfection. Cells treated with 5µg/ml were harvested at 0h, 2h, 4h and 6h while 2.5µg/ml and 10µg/ml treated cells were harvested at 6h post-treatment. The results showed that 2.5µg/ml did not efficiently inhibit mRNA production, while 5µg/ml and 10µg/ml gave similar results. Therefore, treatment with 5µg/ml was used for this experiment. The mRNA half-life was predicted to be between 2h and 3h post treatment. Therefore, the following modifications were made to the conditions; 5µg/ml actinomycin D was used, and cells were harvested at 0h, 2h, 3h and 4h post treatment.

The four plasmids (WT, c.2385T>C, c.2365A>G and both c.2385T>C and c.2365A>G) were transfected into HEK293T cells and treated with 5µg/ml actinomycin D and mRNA was harvested. The results indicated that both variants significantly increased RNA half-life compared to the WT (Figure 3-15, 3-16 and 3-17). In addition, the presence of both variants together in the same plasmid showed a combined increase in mRNA half-life suggesting that each of these two variants increase mRNA half-life resulting in the observed increase in the mRNA and levels of protein expression.

In order to test how these SNV influence mRNA half-life, their effect on mRNA secondary structure was tested using the *in silico* RNAsnp (Section 2.2.1). However, they were not predicted to have any significant effect (c.2365A>G p=0.57 and c.2385T>C p=0.22). Given that miRNA has been suggested to

influence mRNA stability (Valencia-Sanchez, *et al* 2006, Fabian, *et al* 2010), the effect of both SNV on miRNA, using miRBASE (Section 2.2.1), was also investigated but again showed that neither had any effect.

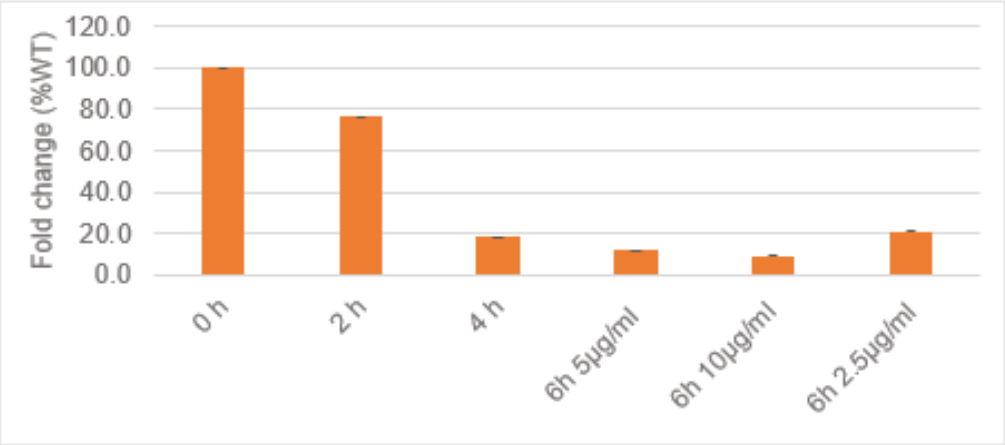


Figure 3-15: RNA half-life for *VWF*. WT at 0h (100%), 2h (76.6%), 4h (18.5%), 6h (11.9%) using 5µg/ml actinomycin D, 6h 10µg/ml (9.4%).and 6h using 2.5µg/ml (21%).

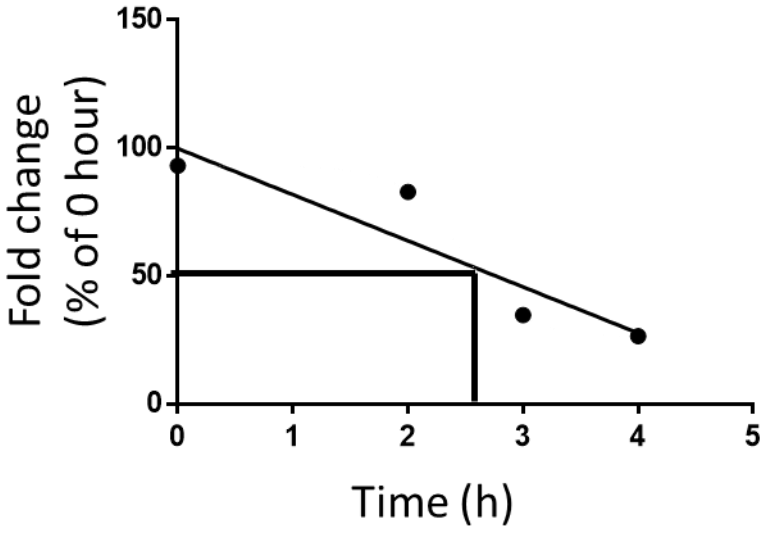


Figure 3-16: Half-life of WT VWF mRNA following actinomycin D treatment in HEK293T cells. Graph used to calculate the half-life of WT VWF mRNA.

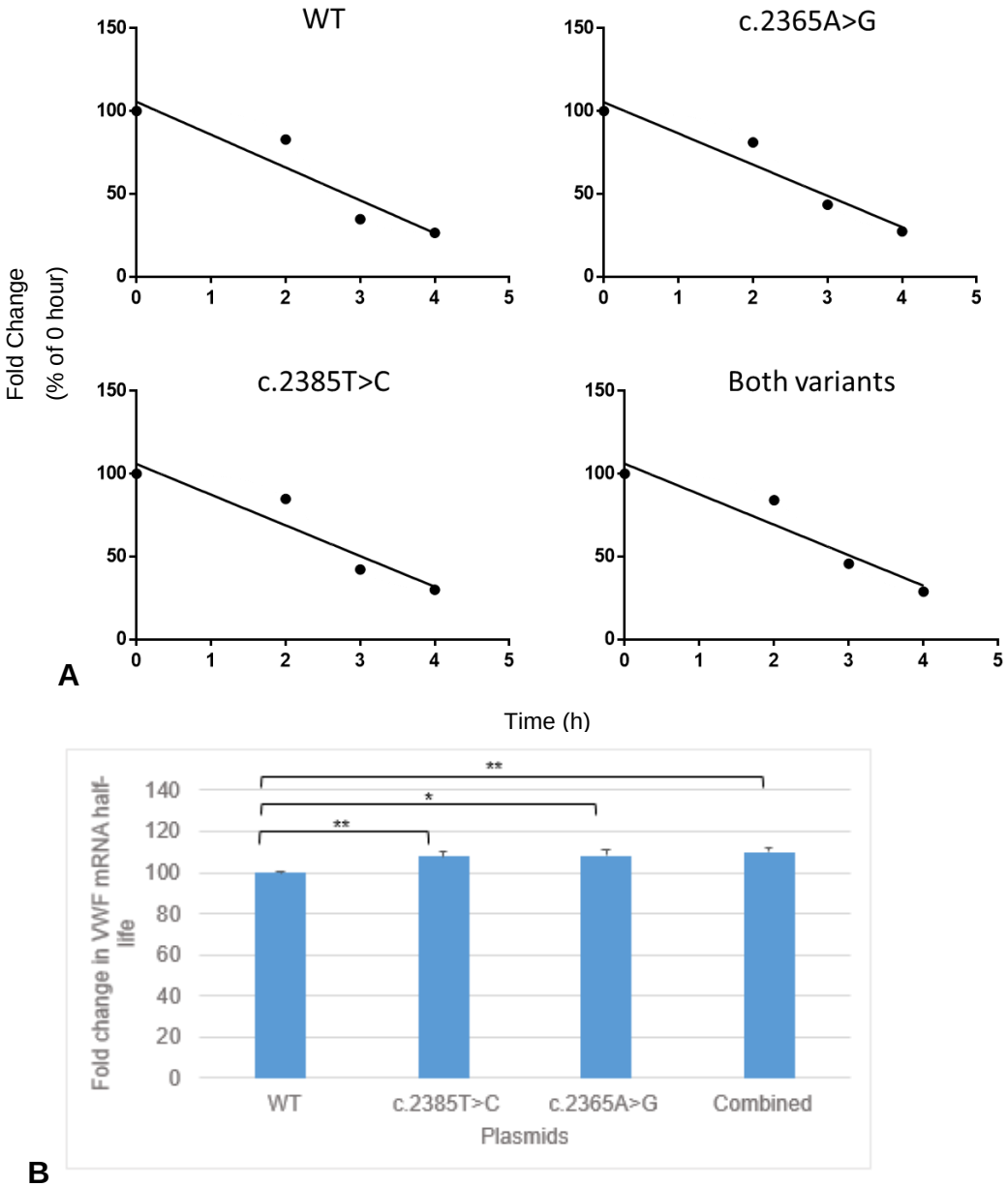


Figure 3-17: mRNA half-life for c.2385T>C, c.2365A>G and both c.2385T>C and c.2365A>G in comparison with WT. A) mRNA half-life was calculated for WT VWF (red), VWF plasmid with c.2385T>C (black), VWF plasmid with c.2365A>G (blue) and VWF plasmid with both c.2385T>C and c.2365A>G *in cis* (green). **B)** mRNA half-life for WT VWF: 100% (2:36 h); c.2385T>C: 107.67% (2:49 h); c.2365A>G: 108.1% (2:51 h); combined: 109.97% (2:54 h). Significance was observed using the ordinary One-way ANOVA test (p=0.0187). The t-test (Mann-Whitney test) was used to compare between two groups (*≤0.05 and **≤0.005). The experiment was repeated three independent times. The bars represent standard error.

3.7 Discussion

The scope of this part of the project was to investigate the effect of *VWF* SNV and examine their possible effect on VWF plasma level. SNV were previously thought to have minor functional importance. However, there is growing evidence that SNV can have an effect on protein function or expression. Even though they might not cause an actual disease, they might contribute to one.

Analysis of SNV throughout *VWF* and its flanking regions revealed several variants with a significant association with VWF plasma level. *In silico* analysis was conducted which showed no major effect on protein structure or function which is not unexpected given that these variants are not disease causing. Following that, *in vitro* investigation was carried out to assess the effect of c.2385T>C and c.2365A>G on protein and mRNA expression as well as their effect on splicing and mRNA half-life. The results suggest that both SNV cause a significant increase in VWF protein and mRNA expression without affecting mRNA splicing, however, these SNV cause an increase in mRNA half-life.

Although these SNV are not disease causing on their own, the presence of several SNV having a similar effect might exhibit or contribute to a disease like phenotype (for example, the presence of several SNVs each of which reduces VWF level could in turn result in an individual with a VWD-like phenotype). In addition, the presence of SNV co-inherited with a *VWF* mutation could result in a more or less severe phenotype (for example, a VWD patient with a SNV increasing VWF level may have a less severe VWD phenotype compared to a patient with the same mutation without this SNV).

In this study, 111 SNV throughout the *VWF* gene and surrounding region were analysed in ~1100 HC from the MCMDM-1VWD study looking for an association with VWF plasma level. Eleven of these SNV were found to be significantly associated with VWF plasma levels including 4 intronic, 4 exonic (3 silent and 1 synonymous), one 3' of *VWF* and two *ANO2* intronic variants. The majority of *VWF* variants were intronic or silent exonic, although these variants were historically thought to have no or a minor effect. However, growing evidence suggesting that these variants are of no less importance

than non-synonymous exonic variants as shown by a recent study that investigated *VWF* variants and their association with VWF level and only identified intronic or silent exonic variants having a significant association with VWF plasma level (Campos*, et al* 2011). Interestingly none of them overlapped with SNV reported in this study apart from c.2385T>C.

Six SNV showed an inconsistent increase or decrease in VWF level when comparing the reference allele, non-reference allele and heterozygous forms which could be due to partial LD with other SNV and therefore these were not investigated further.

On the other hand, some SNV were significantly associated with VWF level and were in strong LD with another SNV such as c.-1342C>G and c.7888-2310T>C, r^2=99.5%, which were associated with a 15% reduction in VWF plasma level and c.2385T>C and c.2365A>G, r^2=99.1%, which were associated with ~9% increase in VWF level in comparison with the reference allele, supporting previous findings for c.2365A>G (Johnsen*, et al* 2013). The reason for not having 100% LD could be due to genotyping errors that might have occurred. The LD between these SNV suggests either that one of the two SNV is associated with VWF level and the other one is showing association with VWF purely due to the strong LD or that both variants are having an additive effect on VWF level. The last SNV significantly associated with VWF plasma level by this study, c.7888-3501A>G was linked to an 8% reduction.

c.-1342C>G and c.7888-2310T>C have not previously been associated with VWF plasma level. c.-1342C>G is present ~1.5kb downstream of *VWF* where no regulatory elements have been predicted to be present. On the other hand, c.7888-2310T>C is present in intron 47 and was not predicted to have any major effect using any web tool suggesting that the observed association is due to another SNV that is in strong LD with these two variants but not included in this study.

c.7888-3501A>G was also present in intron 47 of VWF within ~1.5kb of c.7888-2310T>C. However, they do not share strong LD between them (r^2=34%).

Investigation of c.7888-3501A>G, c.7888-2310T>C and all surrounding SNV within 2kb each side was carried out using various web tools looking at their possible effect on mRNA splicing, ESE motifs, miRNA binding sites and transcription factor binding sites. c.7888-3501A>G was predicted to introduce a new donor splice site by four out of five web tools. This strongly suggests that this SNV is associated with VWF plasma levels through interfering with the splicing system. However, due to time limitations, the intronic variants were not investigated further.

c.2385T>C and c.2365A>G have previously been strongly associated with VWF plasma level (Smith*, et al* 2010, Zhou*, et al* 2014). In this study, both SNV were associated with increased VWF levels which is in line with a previous report (Zabaneh*, et al* 2011), but unlike anther more recent study that did not identify a significant association probably due to the small sample size (Shahbazi*, et al* 2012). However, their mechanism of action has not been previously investigated. Therefore, further work was conducted to investigate this.

The CHARGE association study found that c.2385T>C had the highest significant association with VWF level (Smith*, et al* 2010), although c.2385T>C and c.2365A>G are in perfect LD and thus expected to have a similar association with VWF plasma level. However, one of the main weaknesses of this study type is that not all variants were analysed in all individuals. In this case c.2385T>C was analysed in more individuals than c.2365A>G, therefore, a higher significance was found with c.2385T>C.

Prediction tools did not show a significant effect for these variants on VWF protein or mRNA splicing. However, performing quantification of protein level in HEK293T cells showed an increase (~40%) in cell lysate for c.2385T>C, ~50% for c.2365A>G and ~90% for both SNV. This suggests that these two

SNV independently cause an increase in VWF level resulting in the observed additive effect.

On the other hand, VWF level in the cell supernatant was found to be similar for all three variants (c.2385T>C, c.2365A>G and both variants together) with ~25% increase in VWF secreted compared to WT. VWF is known to go through extensive post-translational modification and complex storage and secretion processes (Lenting*, et al* 2015), so even though the intracellular VWF level is higher in the cells that have both variants, the cells could not process and release VWF faster than cells with only one of the variants. This means the expression of VWF increased while the cells' processing capability had not changed resulting in intracellular retention (in this case due to inability to process higher quantities of VWF). Therefore, both SNV are predicted to increase VWF expression rate in comparison with WT.

To investigate the effect of these two SNV on mRNA expression, quantification of mRNA was carried out. HEK293T cells were transfected with the same plasmid and procedure used for protein quantification in order to produce consistent results. mRNA expression was found to be increased by 50% for c.2385T>C, 50% for c.2365A>G and 70% for both variants *in cis*. These results support previous findings that these two SNV increase VWF expression rate in comparison to WT and they do so independently resulting in an additive increase in level when present together. These findings suggest that the increase in protein level is due to an increase in mRNA expression.

Although the two SNV were not predicted to have a major effect on splicing using various *in silico* tools, it is known that these tools are not error proof. In addition, given that c.2385T>C results in a silent change, it was still a possibility that one or both SNV had an effect on splicing. Therefore, acceptor / donor motif splicing assays were used to investigate this possibility. The results showed no effect of these two SNV on splicing. Following that, the effect of these SNV on ESE motifs was also investigated but similarly neither of the two SNV showed any effect.

The effect of the two SNV on VWF mRNA half-life was subsequently investigated. The results suggested that c.2385T>C and c.2365A>G increase mRNA half-life by ~7.7% (2:49 h:min) and ~8.1% (2:51 h:min) respectively in comparison with the reference allele. An additive increase was observed when both SNV were present together ~10% (2:54 h:min). These findings suggest that the increase in mRNA and protein levels was due to an increase in mRNA half-life. The reason for increased mRNA half-life is not clear given that these SNV do not have an effect on splicing, mRNA secondary structure or an effect on miRNA. However, a recent study suggested that a nucleic acid change from A or T to C or G (which is the case for both SNV) increases mRNA stability (by a yet to be known mechanism) and thus mRNA half-life (Duan*, et al* 2013).

The current study provides evidence that c.2385T>C and c.2365A>G are associated with an increase of VWF level in the normal population. Both SNV are shown to have an independent and an additive effect on VWF level. The two SNV result in an increase in mRNA half-life which in turn results in an increase in the intracellular mRNA quality, and therefore, a higher VWF protein level. This increase is not expected to be disease causing. However, it can be considered as a risk factor for cardiovascular disease where elevated VWF level is a risk factor such as in venous thrombosis, myocardial infarction or coronary heart disease (Jansson*, et al* 1991, Thompson*, et al* 1995, Lacquemant*, et al* 2000, Morange*, et al* 2011, Morange*, et al* 2015). This data also supports previous findings that c.2385T>C is a risk factor for cardiovascular diseases (van Schie*, et al* 2011a). An earlier study has also suggested an association of this variant with coronary heart disease (Lacquemant*, et al* 2000). In conclusion, these findings support the hypothesis that SNV within *VWF* influence VWF plasma level in the normal population.

4 Chapter four: Association of *VWF* SNV c.2771G>A and c.4146G>C with VWF level

4.1 Background

Previous work within this research group identified two SNV as being associated with VWF plasma level, c.4146G>C (p.Lys1382=) and c.2771G>A (p.Arg924Gln; rs33978901).

Previous studies have identified c.2771G>A (occurring in *VWF* exon 21) and classified it as a VWD causative mutation (Goodeve*, et al* 2007, James*, et al* 2007) while other studies suggested it to be benign (Cumming*, et al* 2006, Lester*, et al* 2008). c.2771G>A was found to be present in HC and index cases (IC) at frequencies of 3.1% and 4.1% in the European population (Table 4-1) and 1.0% and 8.5% in the Canadian population respectively (Goodeve*, et al* 2007, James*, et al* 2007). More recent studies observed a significant reduction in VWF plasma level in the heterozygous state of this variant, yet, *in vitro* expression did not show a significant reduction in VWF level nor an effect on splicing using *in silico* web tools (Berber*, et al* 2009, Hickson*, et al* 2010). However, abnormal mRNA splicing was observed in an individual carrying c.2771G>A that resulted from introduction of a possible cryptic splice site (i.e. where an alternate splice site is recognised due to the presence of sequence variation) (Berber*, et al* 2009).

c.4146G>C is silent change occurring in *VWF* exon 28 (which encodes the A1 domain) and was originally identified in a VWD patient (P10F5II:2, with VWF level = 48 IU/dL) from the MCMDM-1VWD study (Hampshire*, et al* 2010). The variant has also since been reported in two Spanish individuals with VWD1 *in cis* with two other variants (Corrales*, et al* 2012), unlike the previous study where the patient had only c.4146G>C (Hampshire*, et al* 2010). Protein expression of this variant was investigated which showed that it caused a significant reduction (~25%) in secreted VWF level (Alyami, PhD thesis, 2014). Further analysis also showed that the variant had no effect on multimerisation, but *in silico* analysis suggested that it might have a possible effect on an ESE (Alyami, PhD thesis, 2014).

Although both variants (c.4146G>C and c.2771G>A) were suggested to be associated with VWD1, the mechanisms by which these SNV have an effect

on VWF level were not clear. Therefore, the aim of this part of the study was to identify the mechanism by which these variants influence VWF level.

Table 4-1: Phenotype of EU members with c.2771G>A.

Family	Individual	Status*	FVIII:C IU/dL	VWF:Rco IU/dL	VWF:Ag IU/dL	Allele 1	Allele 2
P8 F1	II:1	IC	61	49	49	p.R924Q	
P9 F14	II:2	IC	7	3	3	p.[R924Q+C1927R]	c.1533+1G>T
P9 F14	I:2	AFM	63	58	57	p.[R924Q+C1927R]	
P9 F14	III:1	AFM	25	18	19	p.[R924Q+C1927R]	
P9 F14	I:1	UFM	82	79	88		c.1533+1G>T
P9 F14	II:3	UFM	50	80	57	p.[R924Q+C1927R]	
P7 F3	II:1	IC	15	38	25	p.R924Q	p.R854Q
P7 F3	I:2	AFM	104	88	81	p.R924Q	
P7 F3	I:1	UFM	77	87	75		p.R854Q
P6 F5	II:1	IC	18	7	11	p.R924Q	p.R1205H
P6 F5	III:1	AFM	28	17	15		p.R1205H
P6 F5	II:3	UFM	107	95	131	p.R924Q	
P6 F11	II:1	IC	52	34	40	p.R924Q	c.3675-14G>A
P6 F11	II:4	AFM	71	50	49	p.R924Q	p.R924Q
P6 F11	III:1	AFM	68	52	62	p.R924Q	
P6 F11	II:3	UFM	77	69	64	p.R924Q	
P6 F11	III:2	UFM	96	82	84		c.3675-14G>A
P3 F13	I:1	IC	78	13	42	p.[R924Q+R1315RL]	
P3 F13	II:1	AFM	35	3	18	p.[R924Q+R1315RL]	p.Y1584C

*IC, index case; AFM and UFM, affected and unaffected family member.

4.2 *In silico* analysis

The effect of c.4146G>C on ESE motifs was previously predicted and suggested a possible effect (Alyami, PhD thesis, 2014). The effect of c.2771G>A on ESE motifs was predicted in this study showing a possible effect (Table 4-2).

Table 4-2: *In silico* prediction of the effect of c.2771G>A and c.4146G>C on ESE using two different tools.

	Human Splice Finder	ESE finder
c.2771G>A	+2 / -1	+2 / -0
c.4146G>C	+1 / -1	+0 / -1

*+ refers to the gain of an ESE motif and - refers to the loss of an ESE motif (i.e. human splice finder predicted c.2771G>A to cause gain of two new ESE motifs and to lose of one ESE motif).

4.3 Exon splice enhancer assay

Investigation of the potential effect of c.4146G>C and c.2771G>A on ESE motifs was conducted *in vitro* using an exon splice enhancer assay (Section 2.2.2.2.1).

The ESE assay was used to investigate the WT codon (c.4146G), mutant codon (c.4146C) and the other 2 possible variants that would result in a leucine codons (c.4146A and c.4146T) in order to assess whether these two hypothetical changes would have an effect similar to the WT or mutant codon. pcDNA-Dup (SF2-ASF3x) plasmid was used for this purpose (Section 2.2.2.2.2). WT and mutant were prepared for each variant and transfected into HEK293T cells. 24hr post-transfection, RNA was extracted, reverse transcription performed and amplification of the product took place. The PCR product was then run using gel electrophoresis to check the product size. The results showed a lower band size for all mutant sequences compared to WT, suggesting that these changes either caused a loss of exon splice enhancer motifs and/or a gain of exon splice silencer motifs resulting in the lack of identification of the middle exon leading to a smaller band size (Figure 4-1A).

On the other hand, c.2771G>A had a single band with the non-reference allele (c.2771A) compared to two bands in the WT state (c.2771G), suggesting that the change caused a gain of exon splice enhancer motif(s) and/or a loss of exon splice silencer motifs (Figure 4-1B).

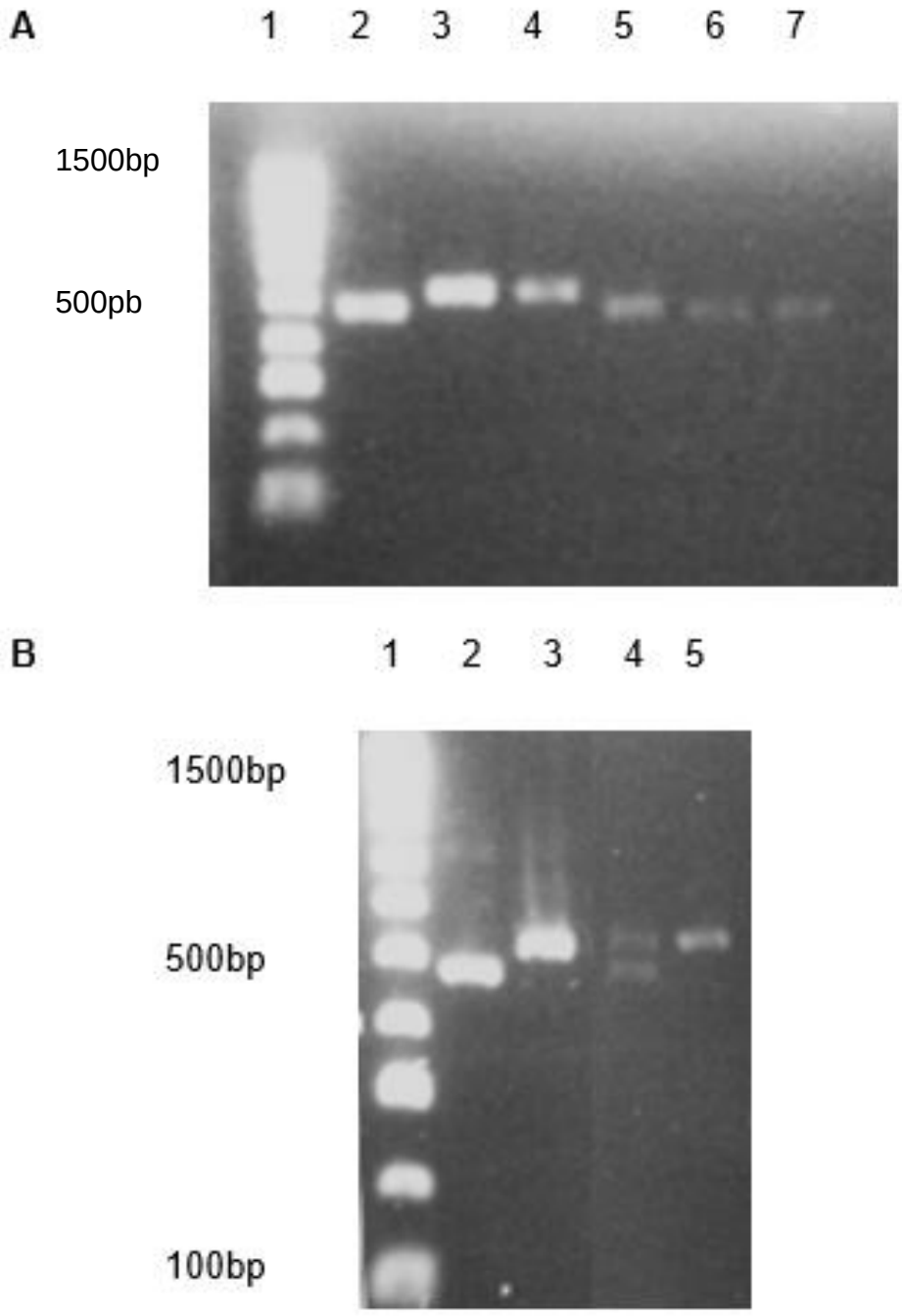


Figure 4-1: The effect of c.4146G>C and c.2771G>A on ESE. A) lane 1: ladder; lane 2: -ve control (470bp); lane 3: +ve control (500bp); lane 4: WT c.4146G; lane 5: mutant c.4146C; lane 6: mutant c.4146A and lane 7: mutant c.4146T. **B)** lane 1- ladder; lane 2- -ve control; lane 3- +ve control; lane 4- WT c.2771G; lane 5- mutant c.2771A.

4.4 RNA quantification

Given that c.4146G>C and c.2771G>A are associated with an alteration in ESE motifs, mRNA level was then investigated in order to assess whether this led to a reduction in mRNA level. This was performed by comparing VWF mRNA expression of both variants with WT. HEK293T cells were transfected with either VWF WT plasmid, VWF c.4146C plasmid or VWF c.2771A plasmid, RNA was extracted 24hr post-transfection followed by reverse transcription and finally mRNA was quantified using qPCR. The results suggested a significant reduction in mRNA level of both variants compared to WT (Figure 4-2).

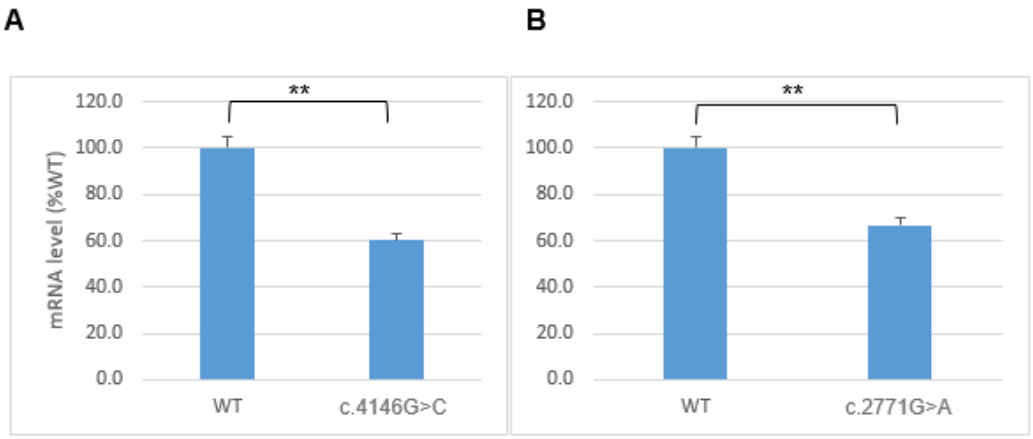


Figure 4-2: VWF mRNA level of c.4146G>C (A) and c.2771G>A (B) compared to WT. Significant differences were observed between WT and c.4146G>C (p=0.001) and between WT and c.2771G>A (p=0.007). Two-tailed Student's t-test. The experiment was done three independent times. Bars indicate standard error.

4.5 RNA degradation rate

The effect of c.4146G>C and c.2771G>A on mRNA half-life was investigated in order to identify whether the alteration in ESE motifs reduced mRNA stability and thus mRNA half-life. Both variants showed a significant reduction (~40%) in mRNA half-life compared to WT (Figure 4-3).

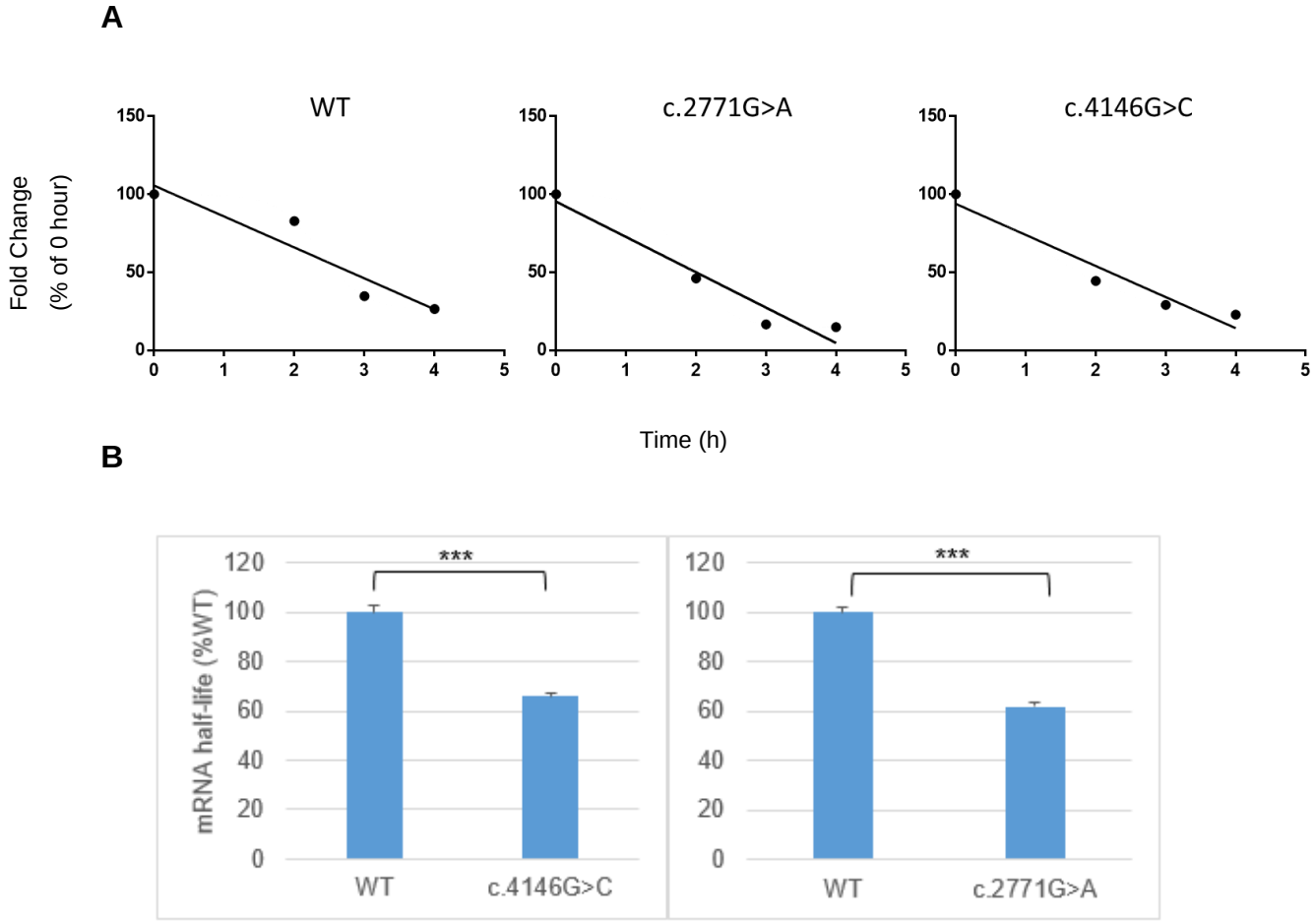


Figure 4-3: mRNA half-life of c.4146G>C and c.2771G>A compared to WT. **A)** mRNA level at 0hr, 2hr, 3hr and 4hr post actinomycin D treatment for WT (2:38 h:min), c.4146G>C (1:44 h:min) and c.2771G>A (1:38 h:min). **B)** Fold change in mRNA half-life for c.4146G>C and c.2771G>A compared to WT. Significant differences were observed between WT and c.4146G>C (p=0.0003) and between WT and c.2771G>A (p=0.0002). Two-tailed Student's t-test. The experiment was done three independent times. Bars indicate standard error.

4.6 Discussion

The aim of this chapter was to investigate the mechanism by which the previously identified variants (c.2771G>A and c.4146G>C) influence VWF plasma level. The two variants were identified in VWD patients and were found to be associated with VWD1. *In vitro* protein expression showed that c.4146G>C was associated with a reduction in secreted VWF level while c.2771G>A showed no effect. However, the previous work did not reach a conclusion on the mechanism of influence of either SNV on VWF level. The current work investigated their possible effect on ESE motifs as well as on mRNA level and mRNA half-life. The current findings suggest that both SNV are associated with an alteration in ESE motifs leading to a defect in splicing, possibly producing an abnormal mRNA that is degraded faster than WT resulting in a reduction in mRNA half-life and mRNA level.

In silico analyses on c.4146G>C and c.2771G>A showed a possible effect on ESE. Therefore, an ESE assay was used to investigate the effect of these SNV on ESE *in vitro*. The results of the ESE splicing assay for the c.4146G>C variant showed that the WT codon (c.4146G) expressed 3 exon product while the mutant codon (c.4146C) and the other 2 possible leucine codons (c.4146A and c.4146T) expressed only 2 exon product suggesting that either the WT codon contained an exon splice enhancer motif(s) required to splice the middle exon or that the other three codons introduce a new exon splice silencer motif(s) resulting in splicing out the middle exon. The effect of the c.4146G>C variant on ESE function appears to explain the VWD1 phenotype as a defect could lead to a splice defect which could result in a premature stop codon.

c.2771G>A is present within the D3 domain of VWF which is important for FVIII binding, multimerisation and storage. However, previous studies showed that this variant had no effect on FVIII binding (Hilbert*, et al* 2003), multimerisation (Hickson*, et al* 2010) or storage (Berber*, et al* 2009). Therefore, the suggestion is that this variant is having an effect at the mRNA level rather than protein level.

Investigating the effect of c.2771G>A on ESE motifs showed that the WT codon expressed both a 2 and a 3 exon product, while the mutant codon expressed only a 3 exon product. This suggest that the mutant codon results in an introduction of a new ESE motif(s) or loss of an exon splice silencer motif(s). This finding is in line with a previous finding where a cryptic splice site was suggested to be the reason for the association of this variant with abnormal *VWF* mRNA (Berber*, et al* 2009). Cryptic splice sites have been observed in many diseases including VWD1 (Gallinaro*, et al* 2006).

However, an older study did not find any defect on splicing in an individual heterozygous for c.2771G>A (Cumming*, et al* 2006) which was explained by Berber*, et al* (2009) as “not all 924Q alleles are created equal”. Nevertheless, a possible explanation could be the presence of another variant within that region that might have a counter effect to c.2771G>A on ESE motifs leading to normal splicing of *VWF*. In addition, haplotype background for 924 alleles can also be responsible for the observed differences.

A similar case was reported where a loss of ESE motif(s) due to a silent exonic variant has recently been associated with factor V deficiency as a result of a splicing defect (Nuzzo*, et al* 2015). The effect of sequence variants on ESE motifs are not commonly reported to be disease causing as they have been under investigated given that they were only recently suggested to be of functional importance suggesting the possibility of overlooking such variants in association with other disorders.

Having observed an effect on ESE motifs, the effect of both c.2771G>A and c.4146G>C on mRNA level was subsequently compared to WT. The results showed that the mRNA level of c.2771G>A was 35% lower compared to WT while the mRNA level of c.4146G>C was 40% lower. The reductions in mRNA levels for both SNV were significant.

mRNA half-life was then compared between the two SNV and WT. The results suggest that both SNV cause a significant reduction in mRNA half-life compared to WT. This could be explained due to an ESE motif defect caused

by both SNV because this could disrupt mRNA stability leading to more rapid RNA degradation.

In summary, both SNV were previously associated with a reduction of VWF level in VWD patients, however, their mechanisms of action had not been previously identified. This study identified that both SNV are associated with defects in ESE motifs which in turn led to a reduction of mRNA stability and thus mRNA half-life. This also suggests that the reduction in mRNA level is due to reduction in mRNA stability instead of a reduction in the level of mRNA expression. Finally, these findings support previous predictions that these two SNV are disease causing.

5 Chapter five: Association of fucosyltransferases *FUT3*, *FUT5* and *FUT6* with VWF plasma level

5.1 Background

There are currently 13 human fucosyltransferases encoding genes reported, *FUT1* to *FUT13* (National Center for Biotechnology Information (NCBI), www.ncbi.nlm.nih.gov, accessed February 2012). Fucosyltransferases add fucose to different glycans and are expressed in different locations (Costache*, et al* 1997). Fucose is added to a basic precursor glycan, converting it to Lewis A antigen (Le^a) either by FUT3 (expressed in exocrine secretions such as saliva) or FUT6 (expressed in plasma), suggesting that FUT6 can possibly substitute for FUT3. Addition of fucose to Le^a by FUT2 (expressed in both exocrine secretions and plasma) further converts it to Lewis B antigen (Le^b) (Figure 5-1) (Mollicone*, et al* 1995, Costache*, et al* 1997).

FUT3 is located on chromosome 19 and has one coding exon of around 1kb (University of California Santa Cruz (UCSC) human genome browser, genome.ucsc.edu, accessed February 2012). The association between *FUT3* and plasma VWF level was first reported by Orstavik and colleagues (1989) who suggested that individuals with Le^a have higher VWF levels compared to those individuals without any Le^a and Le^b, or with Le^b only. A later study showed similar results (Green*, et al* 1995). Recently, these previous observations were further confirmed by research conducted in Sheffield using a larger number of samples, which showed that individuals with inactive FUT3 have ~10% lower VWF plasma levels (Figure 5-2) (Hickson*, et al* 2009). This study focused on a c.202T>C variant in *FUT3* (rs812936) because it was the most common variant in the Caucasian population that results in inactivation of FUT3 as a result of the significant differences of the a.a changed (p.Trp68Arg) (Elmgren*, et al* 1997, Hickson*, et al* 2009, Soejima*, et al* 2009). It was also reported in another study to be one of the common *FUT3* variants leading to its inactivity and Le- (Salomaa*, et al* 2000). However, *FUT3* is reported to be expressed in exocrine secretions but not in plasma, thus its influence on VWF level in plasma is questionable (Costache*, et al* 1997).

The association of *FUT* genes with VWF level has been previously studied including *FUT3*, *FUT1* and *FUT2,* with only *FUT3* showing significant

association (Hickson, PhD thesis, 2010). However, due to the lack of expression of *FUT3* in endothelial cells, other *FUT* genes were investigated.

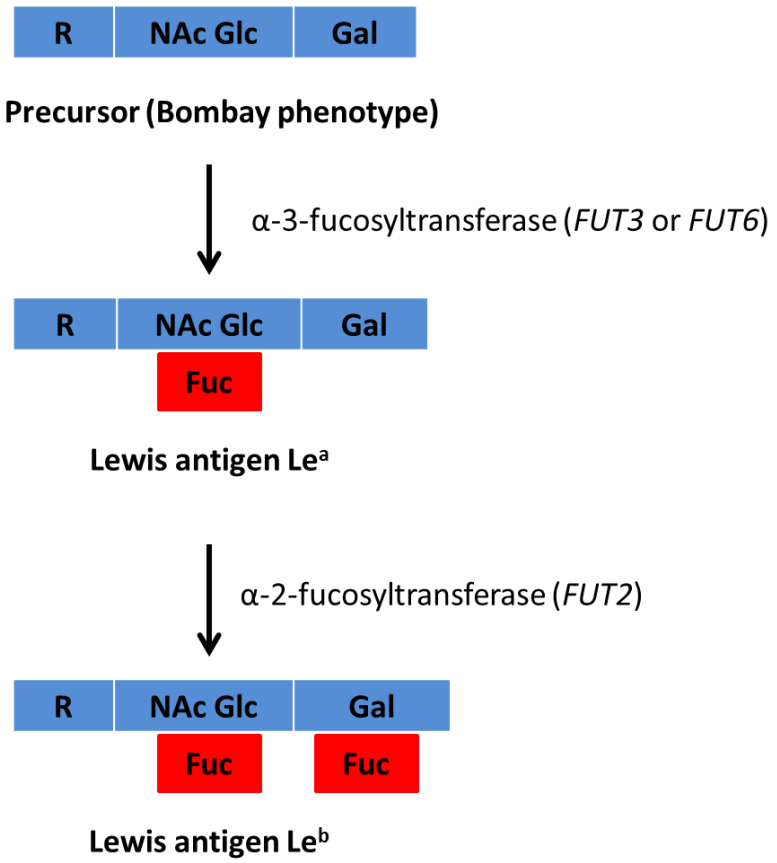


Figure 5-1: Lewis antigen synthesis. The figure shows the process resulting in the conversion of the precursor glycan to Le^a by α-3-fucosyltransferase (encoded by *FUT3* and *FUT6*) and Le^b by α-2-fucosyltransferase (encoded by *FUT2*).

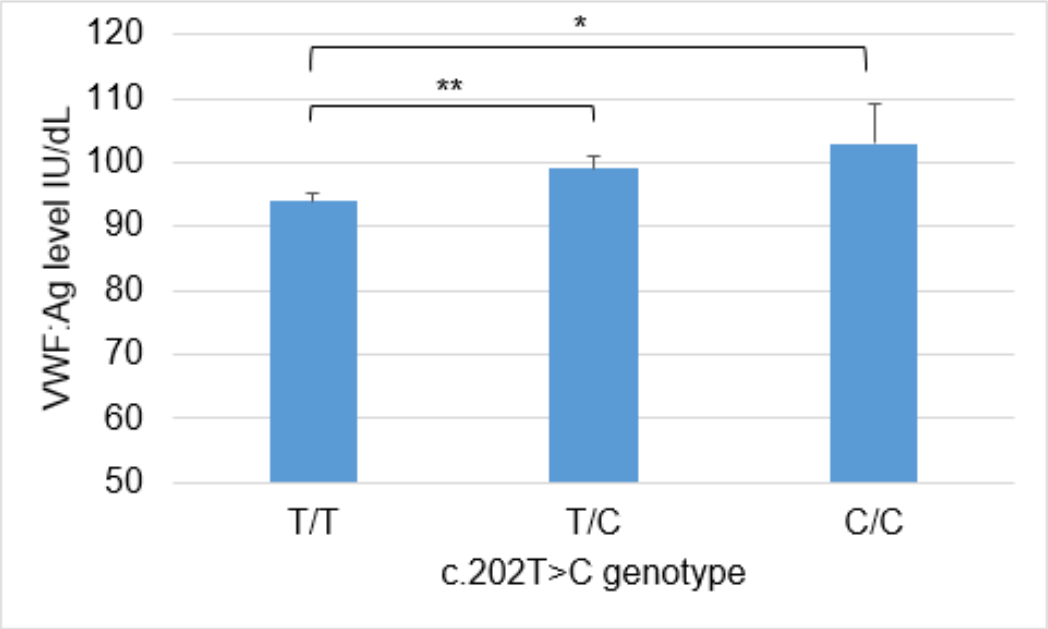


Figure 5-2: Association of SNV c.202T>C in *FUT3* with VWF plasma level in HC. Comparing the three groups showed significant association using the One-way ANOVA (Kruskal-Wallis test) (p=0.01). Comparing T/T with T/C (p=0.0072) and T/T with C/C (p=0.041) was significant using the t-test (Mann-Whitney test), (Hardy-Weinberg p=0.033) (data from (Hickson*, et al* 2009)).

FUT6 is a member of the fucosyltransferase family and has been linked to the conversion of the precursor glycan to Le^a antigen in plasma (Mollicone*, et al* 1995). Given that *FUT6* is expressed in plasma, shares 84% homology with *FUT3* and is located ~5kb from *FUT3* (Figure 5-3) (NCBI, www.ncbi.nlm.nih.gov, accessed February 2012), *FUT6* may be responsible for the observed *FUT3* association. However, whether *FUT6* actually influences VWF plasma level and the mechanism by which fucosyltransferases influence VWF level is not currently known.

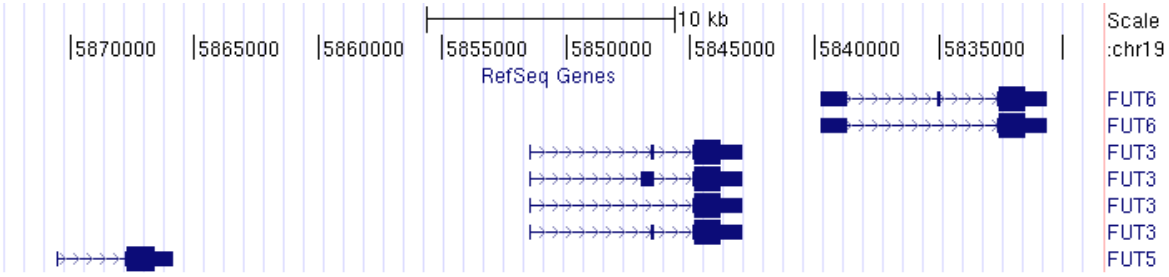
Several studies have investigated mutations influencing FUT6 function or *FUT6* expression. Variants p.Gly739Ala and p.Cys954Ala were found to inactivate the expression of *FUT6* (Mollicone*, et al* 1994, Brinkman-Van der Linden*, et al* 1996, Elmgren*, et al* 2000), while variants p.Cys730Gly and p.Cys907Gly were found to reduce the activity of FUT6 (Pang*, et al* 1999,

Elmgren*, et al* 2000), and variant p.Cys370Thr was found to increase the activity of FUT6 (Mollicone*, et al* 1994, Pang*, et al* 1999, Elmgren*, et al* 2000). However, none of these studies investigated the association of these variants with VWF plasma level.

Based on the close proximity (~60kb) and 85% homology (Figure 5-3) observed between genes *FUT3*, *FUT5*, and *FUT6*, further analysis was conducted to investigate whether any of the three genes were associated with VWF plasma level. A phylogram generated to predict the evolution of *FUT* genes indicates that *FUT5* and *FUT6* are more closely related to *FUT3* than any of the other *FUT* genes (Figure 5-4).

Therefore, the aim was to investigate the association of *FUT3*, *FUT5* and *FUT6* with VWF plasma level. We hypothesised that the apparent association of the *FUT3* SNV to VWF levels is influenced by variation in *FUT5* and/or *FUT6* with which the *FUT3* SNV is in strong LD.

A



B

```
FUT3          1 GSSRQD--TTPTRPTLLILLWTWPFHIPVALSRCSEMVPGTADCHITADRKVYPQADTVI
FUT5          1 GSRCQDSMATPAHPTLLILLWTWPFNTPVALPRCSEMVPGAADCNITADSSVYPQADAVI
FUT6          1 GSRFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVYPQADAVI
consensus     1 **   *   **.   *********  *.** ******** *** ****  ****** **

FUT3         59 VHHWDIMSNPKSRLPPSPRPQGQRWIWFNLEPPPNCQHLEALDRYFNLTMSYRSDSDIFT
FUT5         61 VHHWDIMYNPSANLPPPTRPQGQRWIWFSMESPSNCRHLEALDGYFNLTMSYRSDSDIFT
FUT6         61 VHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSHCWQLKAMDGYFNLTMSYRSDSDIFT
consensus    61 *** ..* **   **   * ******** .* *  *  * *.* ****************

FUT3        119 PYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWKPDSARVRYYQSLQAHLKVDVYGRSHK
FUT5        121 PYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWKPDSARVRYYQSLQAHLKVDVYGRSHK
FUT6        121 PYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKVDVYGRSHK
consensus   121 ********************************** * ************************

FUT3        179 PLPKGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPSRSNYERFLPP
FUT5        181 PLPKGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPSRSNYERFLPP
FUT6        181 PLPQGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPSRRNLIFLGPH
consensus   181 *** ********************************************** *     *

FUT3        239 DAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWALDFCKACWKLQQE
FUT5        241 DAFIHVDDFQSPKDLARYLQELDKDHARYLSYFHWRETLRPRSFSWALAFCKACWKLQQE
FUT6        241 LSGGLIYLRTRLPEASPA------------------------------------------
consensus   241      .       .

FUT3        299 SRYQTVRSIAAWFT
FUT5        301 SRYQTVRSIAAWFT
FUT6            --------------
consensus   301
```

Figure 5-3: The *FUT3*, *FUT5* and *FUT6* locus on chromosome 19. The figure shows: A) the location of (from left to right) *FUT5*, *FUT3* and *FUT6* on chromosome 19. B) an alignment between FUT3, FUT5 and FUT6 proteins and shows the similarity between their a.a sequences. Completely conserved a.a are highlighted in black, while a.a highlighted grey have similar chemical characteristics.

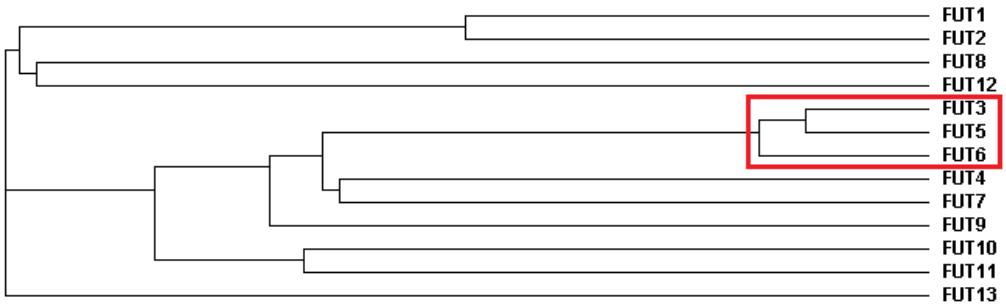


Figure 5-4: Predicted phylogram of the evolution of *FUT* genes.
Generated from CLUSTALW2 phylogram prediction.

5.2 In silico analysis of *FUT3*, *FUT5* and *FUT6*

The linkage disequilibrium (LD) between c.202T>C and other variants in *FUT5* and *FUT6* was investigated based on the theory that the observed effect of c.202T>C on VWF level was actually due to another variant in strong LD with it.

Using HapMap data (accessed August 2013) there was a high degree of LD observed between variants within *FUT3* and *FUT6*, while *FUT5* was not in strong LD with either *FUT3* or *FUT6* (Figure 5-5). The coding region including 100 bases flanking each of three genes was amplified and sequenced in ~100 healthy control individuals (HC) and LD identified between variants in the three genes was investigated using SNPStats (Figure 5-6). The investigation resulted in identification of two SNVs in *FUT6* which were in strong LD with c.202T>C, rs778805 (c.370C>T; p.Pro124Ser) and rs61739552 (c.977G>A; p.Arg326Gln).

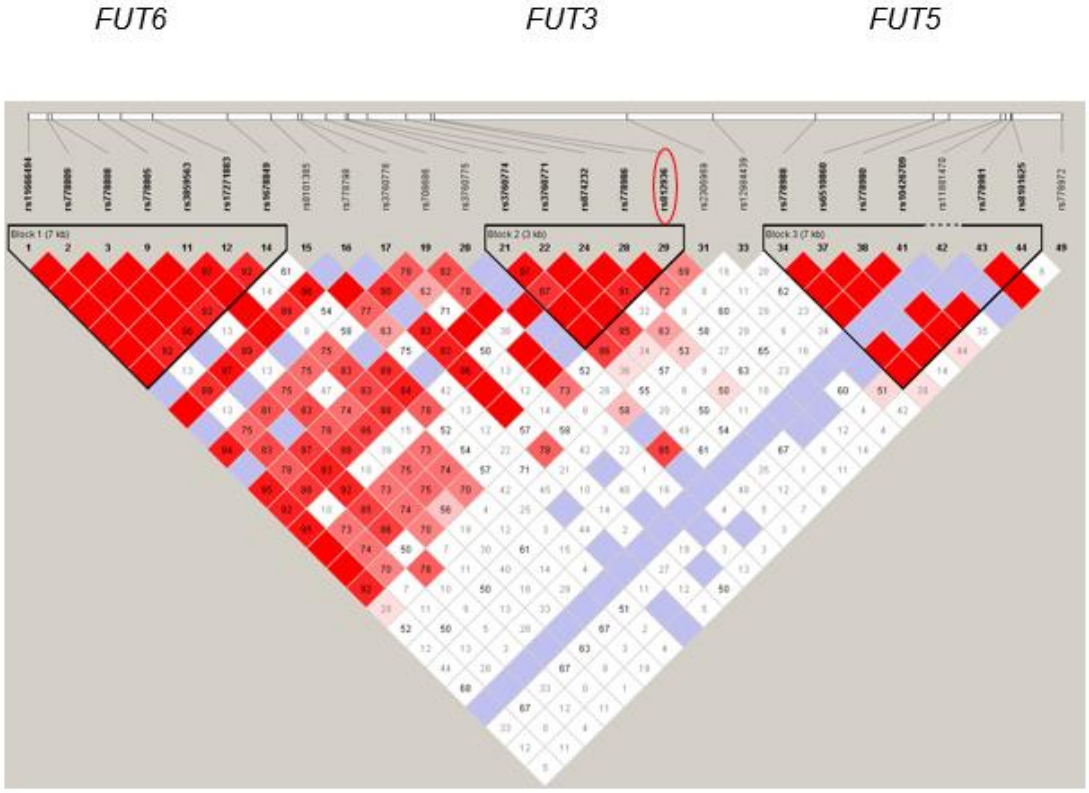


Figure 5-5: Observed LD between *FUT3*, *FUT5* and *FUT6* based on the available HapMap data. White squares refer to weak LD, red squares refer to strong LD while blue squares refer to missing data. The number represents the percentage of the LD. Highlighted in a red circle is the previously identified *FUT3* variant c.202T>C.

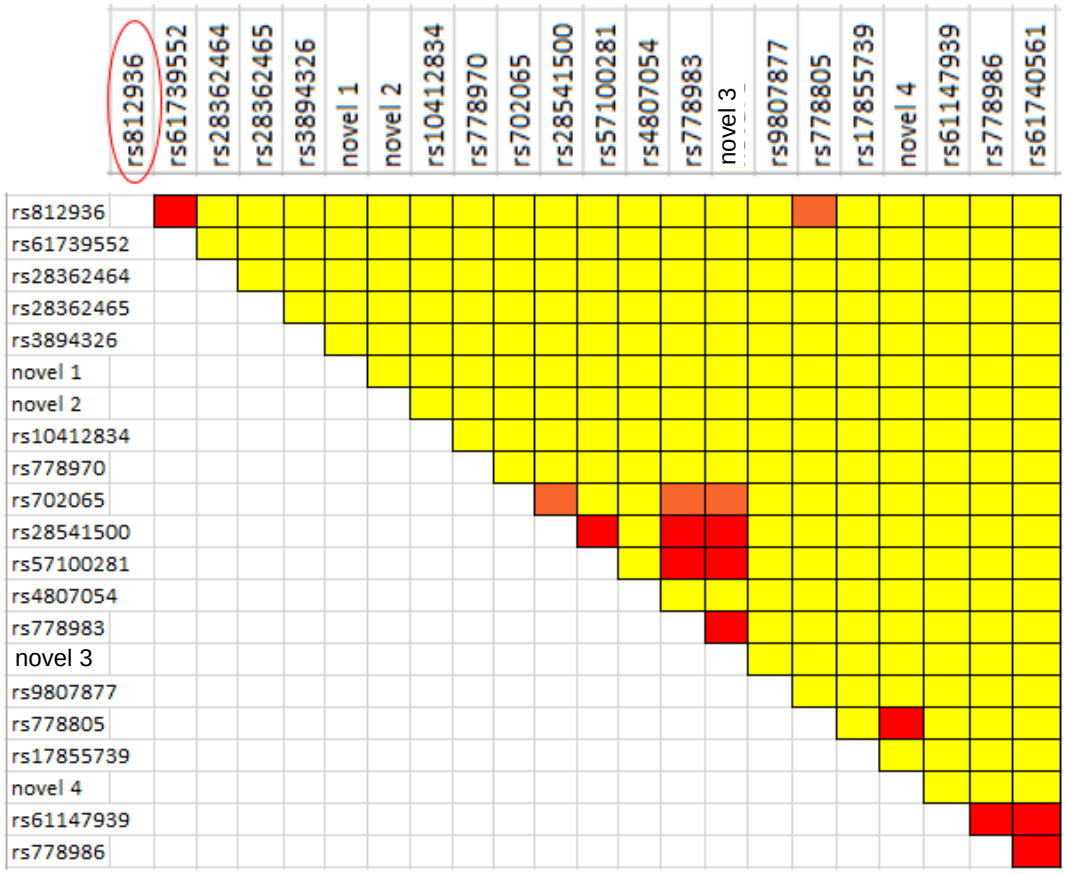


Figure 5-6: LD between reported SNVs in *FUT3*, *FUT5* and *FUT6* when analysed in 96 HC. Yellow indicates very weak LD (<50%); orange indicates medium LD (50-75%); red indicates strong LD (>75%). “Novel” are variants found in our population that were not previously reported. The *FUT3* variant (c.202T>C) is highlighted with a red oval.

5.3 Prediction of *FUT6* variant effect on the FUT6 protein

Following the identification of variants in *FUT6* with a possible effect on VWF plasma level due to their strong LD with the previously identified c.202T>C variant in *FUT3*, these two variants were tested for any effect on protein structure or function using different online prediction tools. One of the three tools predicted these two variants to have an impact on the FUT6 protein and therefore, a potential effect on VWF plasma level even though these tools are not always reliable, instead they give an indication (Table 5-1).

Table 5-1: Prediction of the effect of c.370C>T and c.977G>A on the FUT6 protein.

FUT6 variants	PolyPhen **	SIFT**	Align GVGD***
c.370C>T	Benign (0.006)	Tolerated (0.15)	Most likely damaging (C65)
c.977G>A	Benign (0.03)	Tolerated (0.33)	Moderate (C35)

*The number between brackets are the scores given by each web tool.

** Score ranges between 0-1 where 1 more likely to have an effect

*** Score can be 0-15-25-35-45-55-65 where higher value is more likely damaging

5.4 Association of *FUT6* SNVs with VWF level

Following TaqMan® genotyping of the two SNV in *FUT6* (c.370C>T and c.977G>A) against the rest of the HC sample population (n=1098), statistical analysis was performed to investigate the association of these two SNVs with VWF plasma level compared to what was previously observed for c.202T>C in *FUT3.* Analysis indicated that neither of the two SNVs were associated with VWF plasma level (Figure 5-7) because these variants were not found to be in strong LD after looking at larger sample size.

This observation could have been caused by a possible genotyping error, therefore, Hardy-Weinberg equilibrium (HWE) was investigated. HWE refers to each allele's frequencies within a population which can be calculated in order to predict possible genotyping errors; it was calculated using SNPStats. HWE was significant for c.202T>C variant in *FUT3* suggesting possible genotyping error (i.e. a missing allele, possibly due to the presence of another SNV within a primer binding site), while both *FUT6* SNV were in HWE.

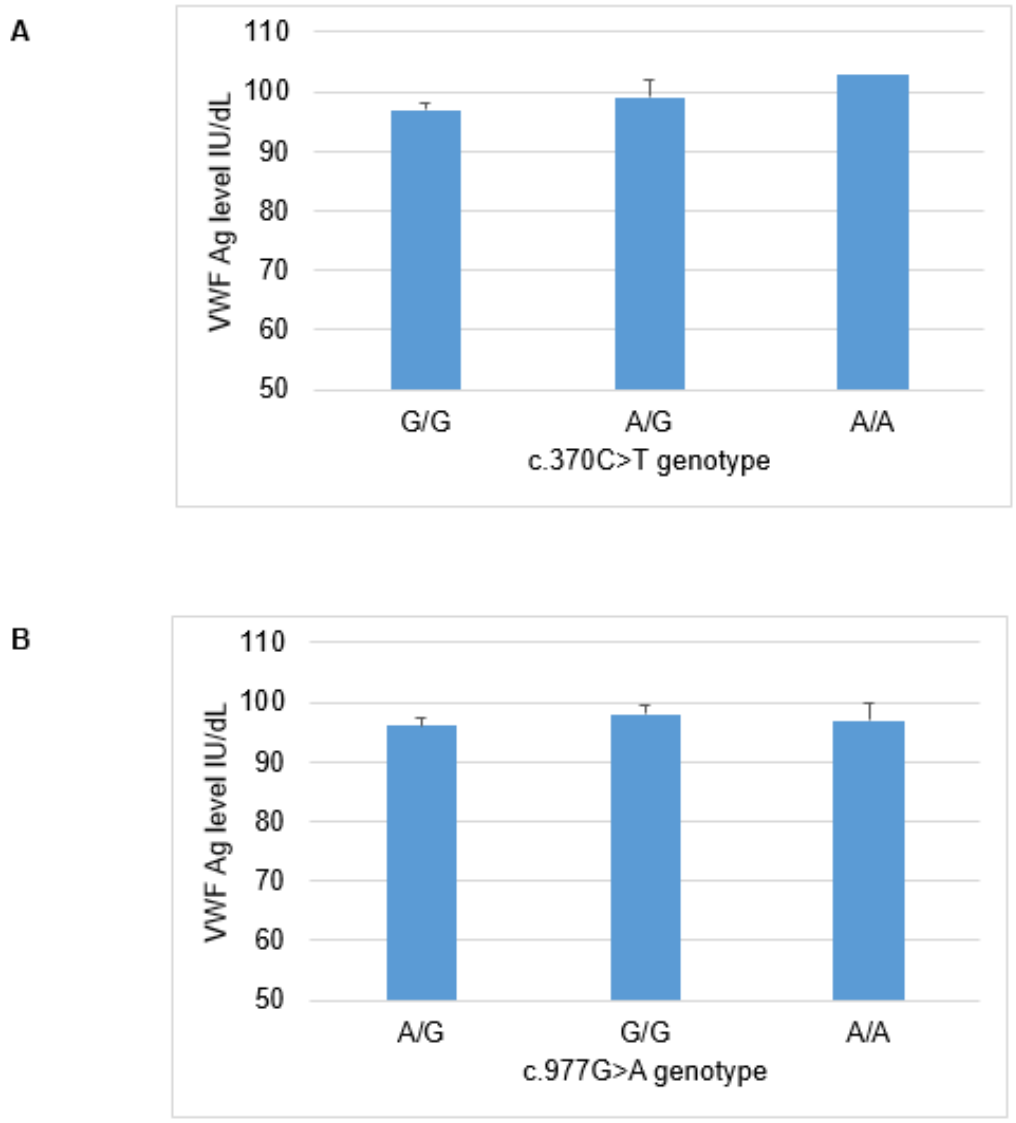


Figure 5-7: Association of c.370C>T and c.977G>A with VWF plasma level. A) SNV c.370C>T in *FUT6* with VWF plasma level in HC (Hardy-Weinberg p=0.63) and B) SNV c.977G>A in *FUT6* with VWF plasma level in HC (Hardy-Weinberg p=0.72). No significant association was observed by comparing the three groups using One-way ANOVA (Kruskal-Wallis test) or by comparing between each groups using the t-test (Mann-Whitney test).

5.5 Re-analysis of LD between SNVs in *FUT3*, *FUT5* and *FUT6*

The LD between *FUT3*, *FUT5* and *FUT6* was reanalysed in an additional 47 HC (to the original 96) with the aim to identify any other SNV(s) in strong LD with c.202T>C. However, the results suggested that no other SNVs were in strong LD with c.202T>C including the previously identified SNVs in *FUT6*. This in turn suggests that c.202T>C in *FUT3* was completely responsible for the observed association with VWF plasma level.

5.6 Association between *FUT3* and VWF clearance

Given that FUT3 is predicted to affect VWF plasma level through influencing glycosylation and thus clearance of VWF, the association between c.202T>C and VWFpp/VWF:Ag ratio was investigated. VWFpp ratio data was available for ~370 HC. Although not significant, the results showed a trend where more clearance (high VWFpp/VWF:Ag ratio) occurred when FUT3 was active (had the reference c.202T allele) (Figure 5-8).

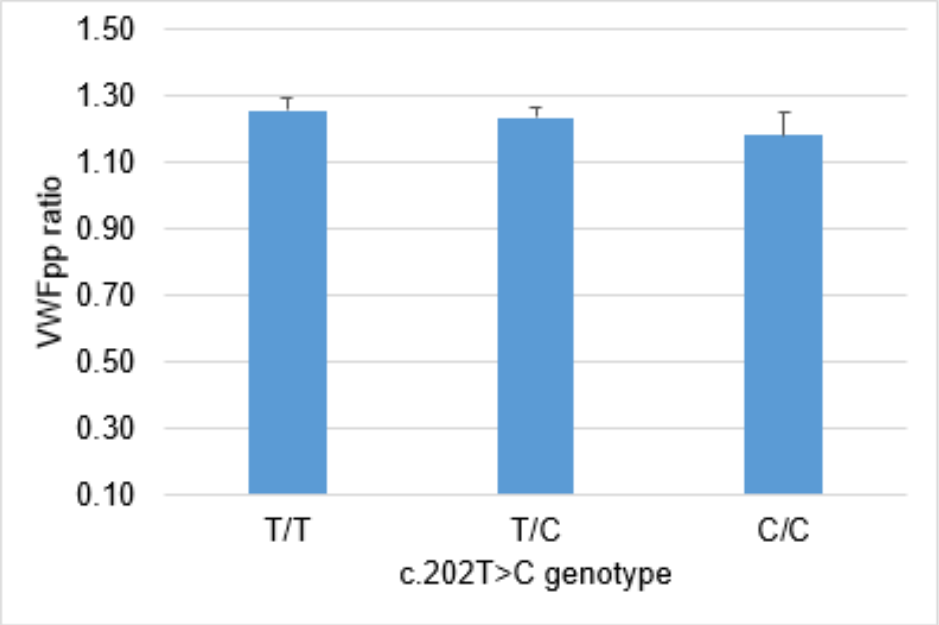


Figure 5-8: Association between c.202T>C and VWFpp/VWF:Ag ratio (p=0.48). No significant association was observed by comparing the three groups using One-way ANOVA (Kruskal-Wallis) or by comparing between each group using the t-test (Mann-Whitney test).

5.7 Investigation of *FUT3* expression in endothelial cells

Given that only *FUT3* showed significant association with VWF plasma level and the lack of evidence of it being expressed in endothelial cells, investigation of the expression of *FUT3* in human umbilical vein endothelial cells (HUVECs) was conducted, given that this is the only endothelial cells that we have access to that normally produce VWF. mRNA (to assess the expression level) as well as DNA (used as positive control) were extracted from HUVECs and qPCR was performed. The same probe was used to detect *FUT3* mRNA and cDNA to avoid any possible differences.

Taking into account that *FUT3* expression might be low, specific primers were designed and the cDNA and DNA were amplified using these primers. *VWF* mRNA was not detected even after (x) cycle of amplification while *VWF* DNA was detected (~33rd cycle) and was detected earlier (~24th cycle) when pre-amplified with specific primers (Figure 5-9). The results suggested that *FUT3* was not expressed in HUVECs. *FUT3* mRNA and DNA were also electrophoresed to confirm the lack of expression obtained from qPCR. *FUT3* DNA could be seen, while *FUT3* mRNA was not detectable.

Following that, *FUT3* mRNA quality and presence were measured using a Nanodrop which confirmed that the RNA obtained was of high quality and quantity. *FUT3* mRNA was also run on RNAse-free gel electrophoresis (Figure 5-10). The results suggested the presence of mRNA and the presence of 18S and 28S bands without any smears again suggesting that the mRNA was of good quality (Figure 5-11). The bands were small which was due to a small volume being used as a limited quantity of mRNA was available.

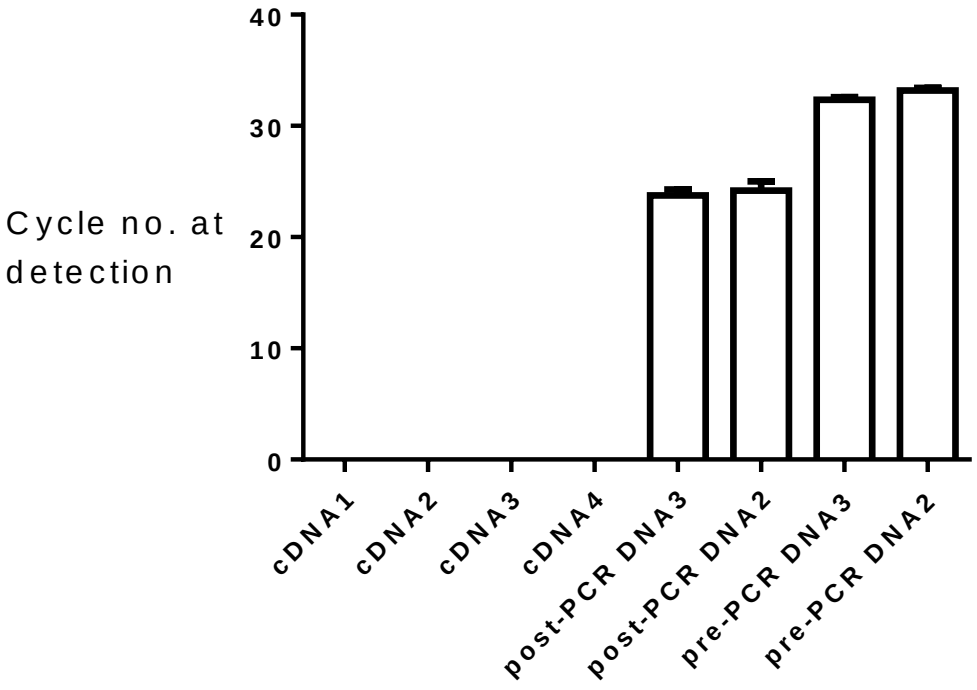


Figure 5-9: *FUT3* mRNA expression in HUVECs. *FUT3* mRNA was not detected even after 40 cycles while DNA was detected at ~35 cycles and DNA amplified by specific primers prior to doing qPCR was detected at ~25 cycles. cDNA was prepared from 4 different individuals.

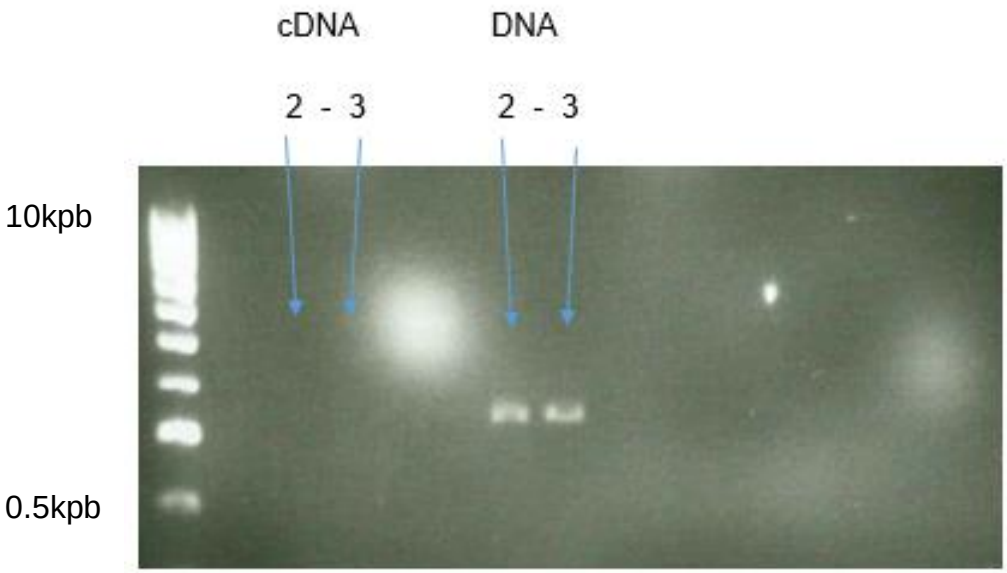


Figure 5-10: Investigation of *FUT3* mRNA expression in HUVECs using PCR and gel electrophoresis. No mRNA was detected in two samples, while DNA was detected in the same individuals.

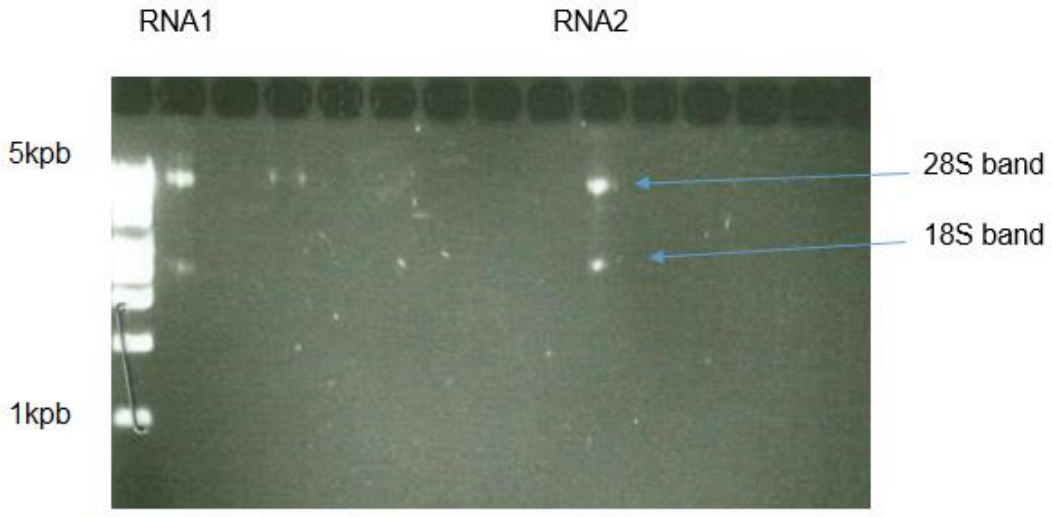


Figure 5-11: The quality of mRNA extracted from HUVECs using gel electrophoresis. 28S and 18S bands were detected in the two mRNA samples.

5.8 Discussion

FUT3 c.202T>C was reported to be associated with VWF plasma level by previous work conducted within this research group (Hickson*, et al* 2009). However, due to the lack of evidence for the expression of *FUT3* in plasma, other genes have also been investigated previously, including *FUT1* and *FUT2*. However, only this variant in *FUT3* seemed to be associated with VWF plasma level. Therefore, this research was conducted to investigate the association of other *FUT* genes including *FUT5* and *FUT6* on VWF plasma level. However, neither of these two genes were associated with VWF plasma level, again supporting the observation that only *FUT3* seemed to be significantly associated with VWF plasma level.

We made the prediction that *FUT3* influences VWF plasma level in similar way as ABO due to the similarity of their function, i.e. via glycosylation and therefore clearance. Given that ABH antigens only account for 13% of the glycans glycosylated on VWF (Matsui*, et al* 1992, McKinnon*, et al* 2008), yet they are associated with ~25% variation in VWF level (Gill*, et al* 1987, McGrath*, et al* 2010). Therefore, other glycans are expected to have an influence on VWF plasma levels.

FUT3 (as well as FUT6) is reported to be required for the conversion of the precursor to Le^a which is then converted to Le^b by FUT2. Therefore, the lack of FUT3 expression due to c.202T>C, leads to lack of both Le^a and Le^b, which might lead to reduction of VWF level similar to that reported for O-blood group (where there is no A or B antigens). However, FUT3 has only been reported to be expressed in exocrine secretions while FUT6 performs a similar function to FUT3 in plasma (Mollicone*, et al* 1994). Therefore, whether FUT3 or FUT6 were responsible for the observed association in HC needed further investigation.

In silico analysis suggested *FUT5* and *FUT6* to be good candidates for further analysis due to their close proximity to *FUT3* and they were also the most related *FUT* genes to *FUT3*. Reinvestigation of *FUT3* was also carried out to

ensure that only the c.202T>C variant in *FUT3* was significantly associated with VWF plasma level.

In subset of random 47 HC samples. LD analysis identified two SNV in *FUT6* that showed strong LD with *FUT3* c.202T>C. However, the results from a larger sample (n=1098) suggested that neither of these SNV was significantly associated with VWF level. This was due to the LD not being strong with c.202T>C after adding more samples.

Sequencing of additional HC for *FUT3*, *FUT5* and *FUT6* did not provide any evidence of further SNV being in strong LD with c.202T>C in *FUT3*. This finding suggests that c.202T>C variant is responsible for the observed association with VWF level instead of other SNV and other *FUT* genes.

To confirm what is reported about the lack of FUT3 expression in plasma, mRNA was extracted from HUVECs, and qPCR was run along with gel electrophoresis. The results did not show the presence of mRNA in these cells confirming previous findings (Costache*, et al* 1997). This in turn suggest that either there is no expression of FUT3 in HUVECs or the expression is very low to a level that is not detectable via the used method. Even though there was no FUT3 detected, there is not enough evidence denying any interaction between FUT3 and VWF, as FUT3 might still be expressed in other endothelial cell types (for example human sinusoidal endothelial cells).

Investigation of the *FUT3* c.202T>C variant with VWF clearance showed a trend similar to that expected (high VWF clearance for active (WT) FUT3), but the differences were small and did not produce significant results due to the small sample size. This suggests that inclusion of more samples may achieve significance.

This study has provided evidence that *FUT3* is associated with VWF plasma level. Even though the previous CHARGE association study did not identify c.202T>C as a candidate locus with a strong association with VWF or FVIII levels (Smith*, et al* 2010), this is potentially due to the fact that c.202T>C was not analysed in the study. Thus, the results of the CHARGE study cannot preclude any association between c.202T>C and VWF plasma level.

The mechanism by which *FUT3* influences VWF plasma level is still unclear. However, a study has shown that Lewis antigens are not glycosylated on VWF (Canis*, et al* 2012), suggesting that either *FUT3* is influencing VWF indirectly or c.202T>C is in strong LD with another SNV not covered in this study (either variants outside FUT5 and FUT6 or in one of these genes but in a region not investigated such as introns or promoter regions).

An indirect effect of FUT3 on VWF could work in one of three ways. First, FUT3 could be performing another unknown function or second, it could be involved in the glycosylation of other proteins that are involved in the processing, secretion or clearance of VWF (such as ADAMTS13 (Sorvillo*, et al* 2014)). Therefore, a defect in FUT3 because of the c.202T>C variant could still lead to the observed association.

Third, ABH and Lewis antigens have been reported to use the same starting precursor (Mollicone*, et al* 1994, Costache*, et al* 1997), thus, the presence or absence of FUT3 could affect the amount of precursor available to be used by ABH and other antigens and thus have an effect on the glycosylation pattern of VWF, and subsequently VWF level through clearance.

In conclusion, the c.202T>C variant in *FUT3* is significantly associated with VWF plasma level. Analysis was previously conducted on *FUT1* and *FUT2* and the current study looked at *FUT5* and *FUT6*, none of which seems to have any association with VWF plasma level. The exact mechanism by which *FUT3* is influencing VWF is still unclear although it appears to be through clearance. Further investigation is therefore required to clarify the mechanism involved.

6 Chapter six: Investigation of the association between *CLEC4M* with VWF level

6.1 Background

CLEC4M (also known as *CD209L*) is located on chromosome 19 (19p13.2-3) and is ~26kb long. It is expressed in the liver by endothelial cells and is also expressed in lymph nodes (Bashirova*, et al* 2001, Chan*, et al* 2006). The encoded protein L-SIGN, is a transmembrane receptor that is involved in the innate immune system by binding to glycosylated and fucosylated products (Guo*, et al* 2004, Zandberg*, et al* 2012). It is involved in pathogen adhesion and recognition; recognising bacteria (such as *Mycobacterium tuberculosis*), parasites (such as *Leishmania pifanoi)* and viruses (such as HIV-1 and Ebola) (Barreiro*, et al* 2005). L-SIGN also acts as a cell adhesion receptor (Feinberg*, et al* 2005, Yu*, et al* 2009, Li*, et al* 2012). However, another study has reported that L-SIGN does not internalise pathogens, but instead recruits macrophages to clear them (Guo*, et al* 2004). The same study reported that L-SIGN binds to ABH antigen as well as Lewis antigens which are known/predicted to be glycosylated on VWF (Guo*, et al* 2004).

L-SIGN is reported to form a tetramer before being expressed on the cell surface (Mitchell*, et al* 2001). Each L-SIGN monomer consists of three main domains; C-terminal, N-terminal and neck region (Bashirova*, et al* 2001, Li*, et al* 2012). The neck region contains a 23 a.a variable number of tandem repeats (VNTR) region ranging from 4–9 repeats in size (Figure 6-1) (Chan*, et al* 2006). A previous study suggested that, longer VNTR repeats enhance the stabilisation and activity of L-SIGN compared to shorter repeats (Feinberg*, et al* 2005), which increases the flexibility of the L-SIGN tetramer binding sites and therefore the ability of L-SIGN to bind to more than one glycan (Feinberg*, et al* 2005). Another study suggested that homozygous L-SIGN tetramers clear pathogens more efficiently compared to heterozygous L-SIGN tetramers, so homozygous tetramers may be more stable regardless of the VNTR size (Soilleux*, et al* 2000, Chan*, et al* 2006, Khoo*, et al* 2008).

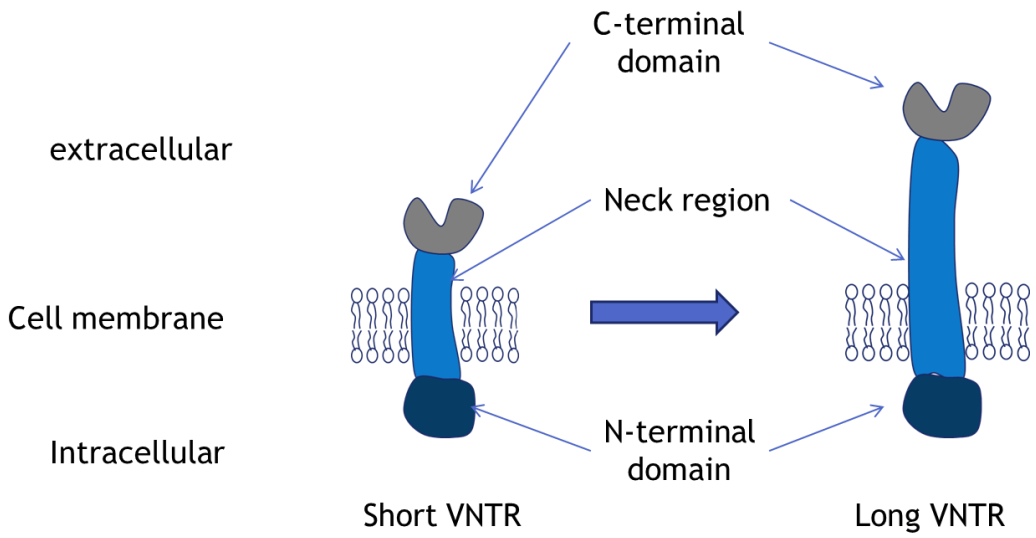


Figure 6-1: Proposed effect of short and long VNTR repeats in the neck region of L-SIGN receptor.

A previous association study identified a variant (rs868875; c.631+73A>G) in *CLEC4M* to be significantly associated with VWF plasma level (Smith, *et al* 2010) which was supported by a recent study in VWD1 patients (Sanders, *et al* 2015). Further work has also suggested an association of the VNTR in the neck region of L-SIGN with VWF plasma level (Rydz, *et al* 2011). A similar observation has previously been reported for the Gplbα receptor where an increase in the neck region size leads to increased exposure of the receptor binding site, which in turn results in increased receptor activity (Lopez, *et al* 1992, Di Paola, *et al* 2005). Because *CLEC4M* is expressed in the liver, it has been suggested that L-SIGN may be involved in the clearance of VWF from the circulation (Rydz, *et al* 2011).

Therefore, the aim of this research was to investigate further the association between *CLEC4M* and VWF levels. We hypothesised that the VNTR in *CLEC4M* is in strong LD with the SNV identified by the CHARGE study (c.631+73A>G) suggesting an association with VWF level through alteration in the neck size.

6.2 *In silico* analysis of *CLEC4M*

CLEC4M was initially investigated to identify any SNV that might be in strong LD with SNV c.631+73A>G, which had been associated with VWF levels by the CHARGE study (Figure 6-2). This was conducted in order to eliminate the possibility that other SNV in strong LD with SNV c.631+73A>G were responsible for the observed association with VWF level. Analysis highlighted that *CLEC4M* was a relatively small gene (compared to other genes highlighted by the CHARGE study) with only a few variants that were in strong LD with SNV c.631+73A>G.

In silico analysis of *CLEC4M* suggested regions with the gene that might be of importance which will be investigated further (Figure 6-3). This includes several SNV found to be in strong LD with the SNV identified in the CHARGE study. In addition, exon 7 of *CLEC4M* was reported to encode amino acids involved in ligand binding, so was also investigated further (Guo*, et al* 2004).

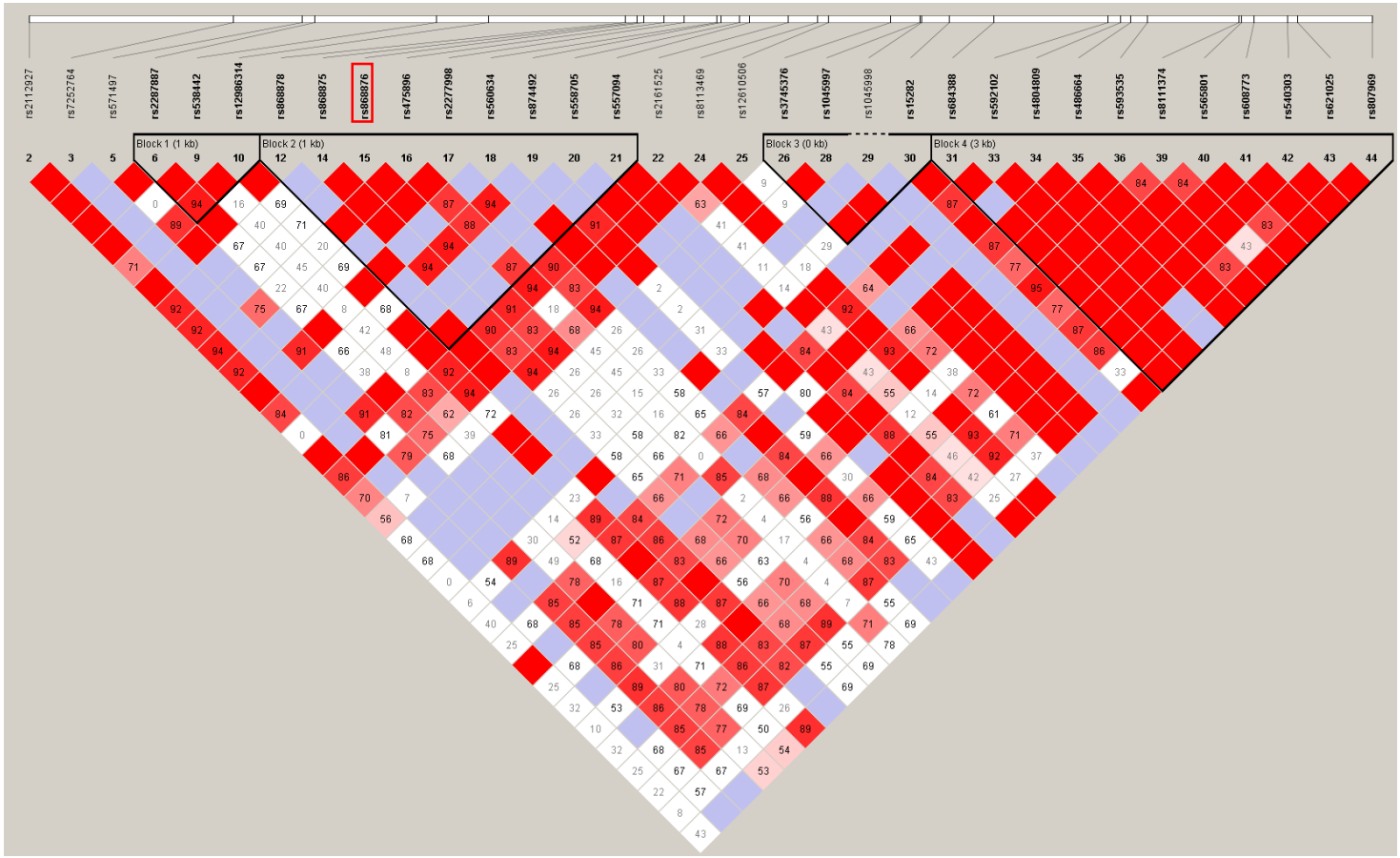


Figure 6-2: LD analysis of the *CLEC4M* locus on chromosome 19. The LD between variants within *CLEC4M* (6.4kb) and ~5kb each side of the gene was investigated. SNV c.631+73A>G identified by the CHARGE association study is highlighted in the red box. White squares indicate weak LD while red squares indicate strong LD, the number in each square shows the percentage of the LD. Grey boxes indicate a lack of data for any given SNV association.

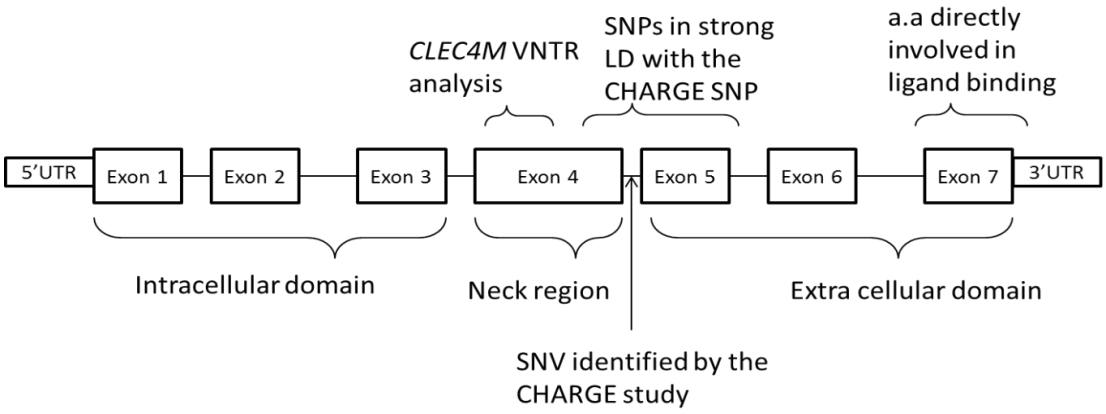


Figure 6-3: The structure of *CLEC4M* highlighting the areas of interest.

6.3 Relationship between *CLEC4M*, *CLEC4L* and *CLEC4G*

CLEC4G, *CLEC4L* and *CLEC4M* are all present within 43kb (chromosome 19) and share homology (*CLEC4M* and *CLEC4L* share 83% homology while *CLEC4M* and *CLEC4G* share 34% homology) (Figure 6-4). All three are also reported to have similar functions but with varying affinities to bind to different sugar groups (Guo*, et al* 2004). The LD between the three genes was investigated by looking for any possible variants in *CLEC4L* or *CLEC4G* that might be in strong LD with the variant identified by the CHARGE study in *CLEC4M*. However, the CHARGE variant did not appear to be in strong LD with any variant in *CLEC4L* or *CLEC4G* (figure 6-4).

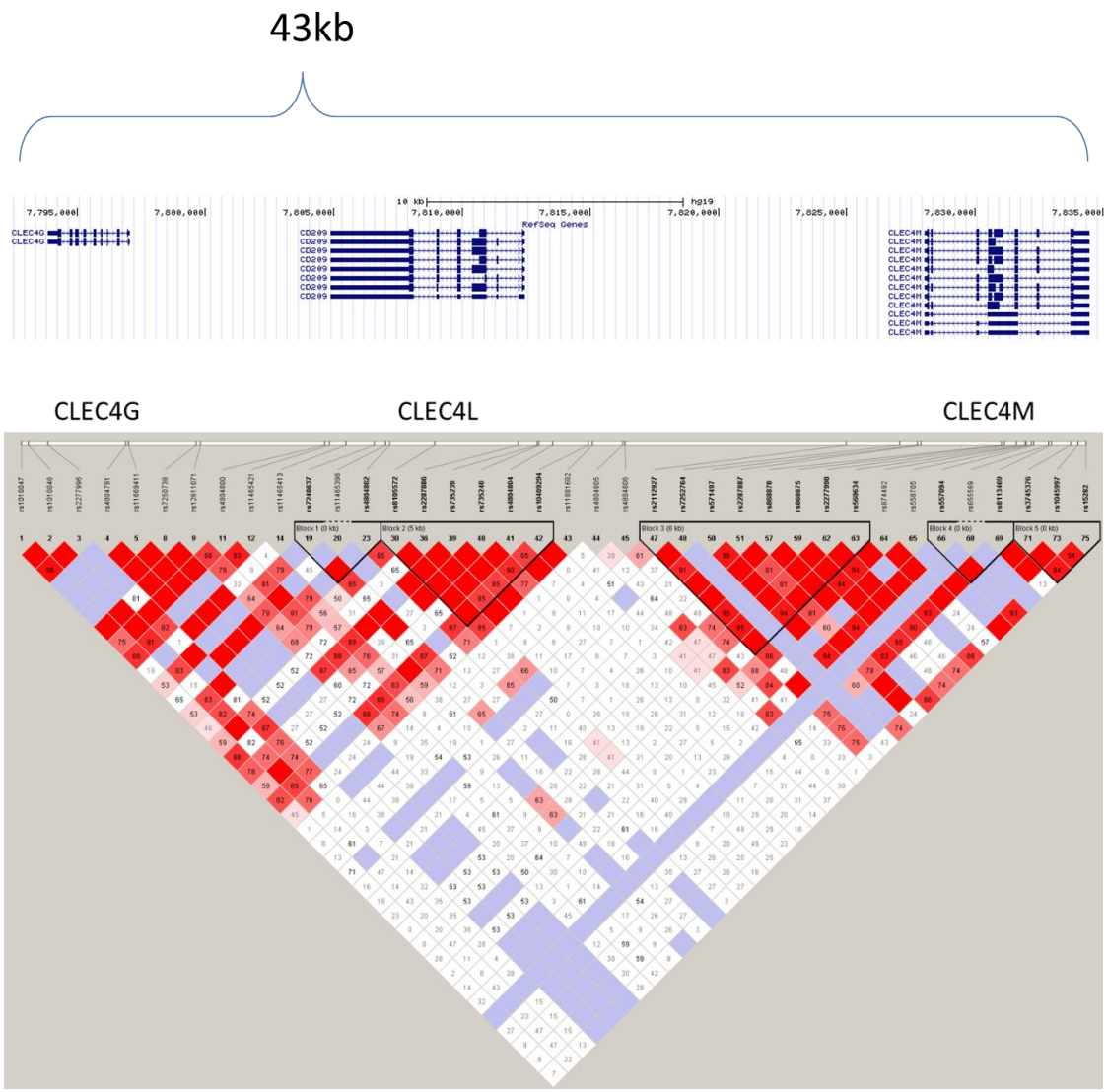


Figure 6-4: The relationship between *CLEC4M*, *CLEC4L* and *CLEC4G*. The location of *CLEC4G*, *CLEC4L* (commonly known as *CD209*) and *CLEC4M* and the LD between *CLEC4G*, *CLEC4L* and *CLEC4M* are shown.

6.4 Investigation of *CLEC4M* variants

2 exonic SNVs were shown to be in strong LD with the CHARGE SNV c.631+73A>G, which are c.718G>A (rs2277998, p.Asp240Asn; NP_001138376) and c.585G>A (rs868878, p.Lys195=, NP_001138376) (Figure 6-5), and were further analysed *in silico* to investigate the possible effect of these SNVs on the function, structure, or expression of L-SIGN and thus a possible effect on VWF level. However, neither of the SNVs were predicted to have a significant effect on *CLEC4M* (Table 6-1).

Further *in silico* investigation was conducted on c.718G>A in exon 5 as it was the only exonic SNV to cause an amino acid change and likely to have an impact on L-SIGN structure or activity. Using *in silico* prediction tools, c.718G>A was shown to have a potential impact on the protein structure, however, c.585G>A was not shown to have any effect and was excluded from any further analysis (Table 6-1). Analysis of the 3D structure of the extra cellular domain of L-SIGN (using PyMol v1.3) showed that SNV c.718G>A was present on the surface (Figure 6-6).

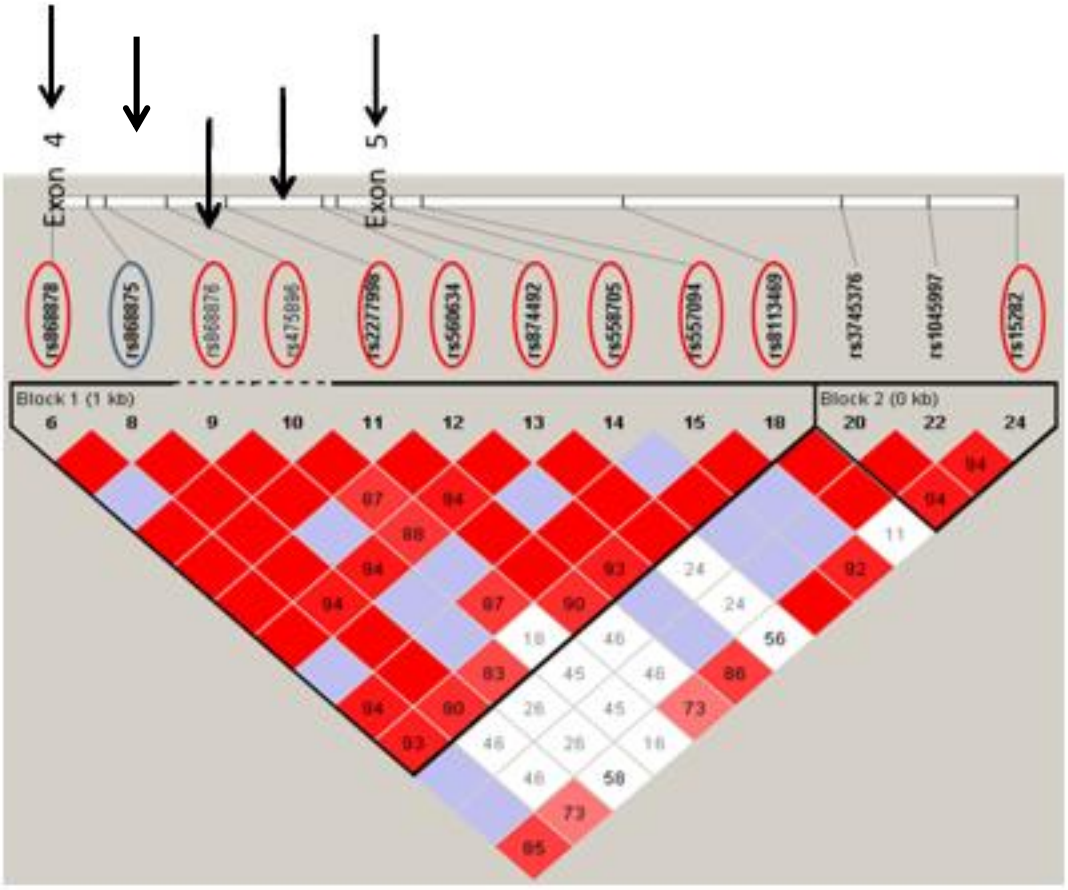
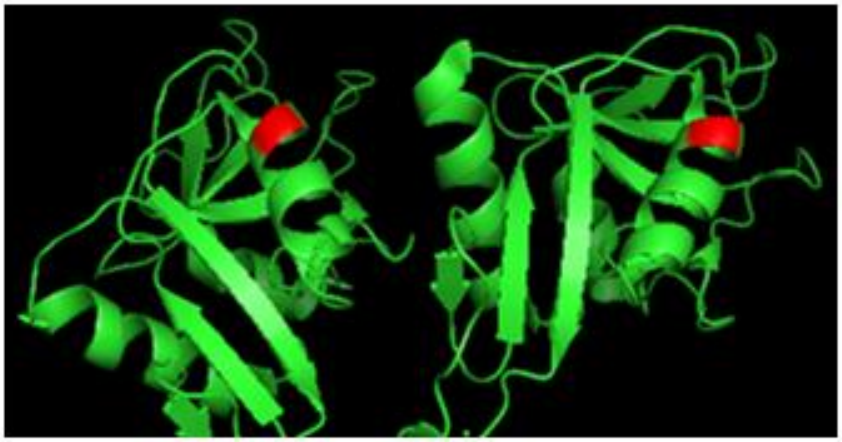


Figure 6-5: LD analysis of the *CLEC4M* SNV identified in the CHARGE study (based on HapMap data). SNVs present in *CLEC4M* exons are indicated. The remaining SNVs were intronic or located in the 3' UTR (rs15282). The numbers within squares refer to the degree of LD (completely red squares represent 100% LD; completely grey squares represent missing data, light red squares refer to weaker LD while white squares refer to lack of LD). The blue oval represents the SNV identified in the CHARGE study. SNVs highlighted in red ovals were in strong LD with the CHARGE SNV. The arrows show the SNVs that are in perfect LD with the CHARGE SNV.

Table 6-1: Prediction of the effect of c.718G>A and c.585G>A on the L-SIGN protein.

CLEC4M variants	SIFT	PolyPhen-2	Align GVGD	Mutation Taster
c.585G>A	Tolerated	Benign	Less likely	Protein structure less likely to be affected
c.718G>A	Tolerated	Benign	Most likely	Might affect protein structure

A)



B)

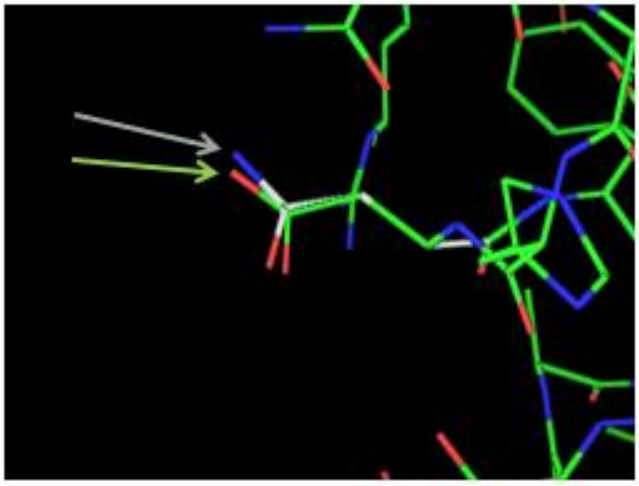


Figure 6-6: 3D structure of the L-SIGN extracellular domain. A) The image represents L-SIGN as a dimer. Highlighted in red is the position of p.Asp240Asn (c.718G>A). **B)** 3D structure of L-SIGN focusing on the position of SNV p.Asp240Asn. Comparison of the WT and mutant alleles at p.240 (green arrow, reference allele; grey arrow, non-reference allele).

Two primer sets were designed to genotype c.718G>A by amplifying and sequencing *CLEC4M* exon 5 in 921 HC from MCMDM-1VWD study. Two sets were designed due to the presence of repetitive sequence within this region in addition to the presence of several different splice variants. The exon was amplified successfully using both primer sets in four randomly selected HC individuals. Primer set 2 was chosen to amplify the rest of the samples because it produced a more consistent amplification (Figure 6-7).

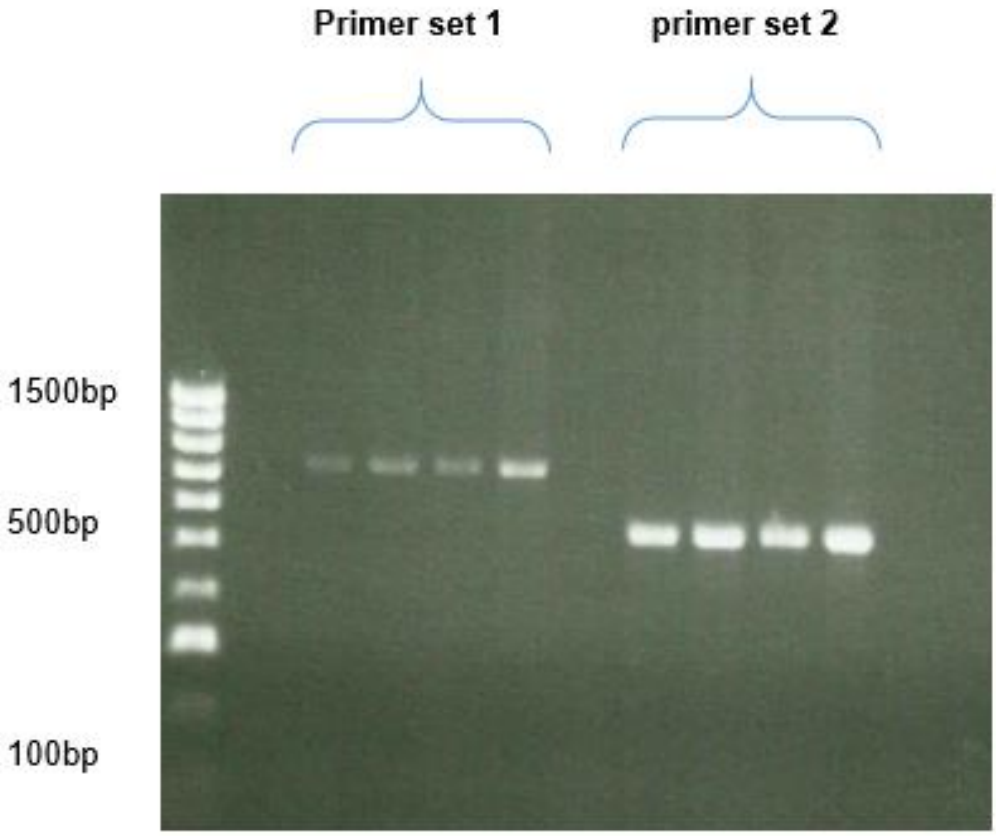


Figure 6-7: PCR product of exon 5 amplified using two different sets of primers in 4 different HC.

After genotyping c.718G>A in the 921 HC, the data showed no significant association with VWF plasma level (Figure 6-8). However, the analysis did show that individuals homozygous for the non-reference allele had a VWF level 5% lower compared to those homozygous for the reference allele ($p>0.05$). Additionally, due to the proximity of SNV c.718G>A and the CHARGE SNV c.631+73A>G, both were genotyped and the data confirmed that both SNV were in strong LD (r^2=98.5%). Both SNV were also shown to be in Hardy-Weinberg equilibrium in the HC samples.

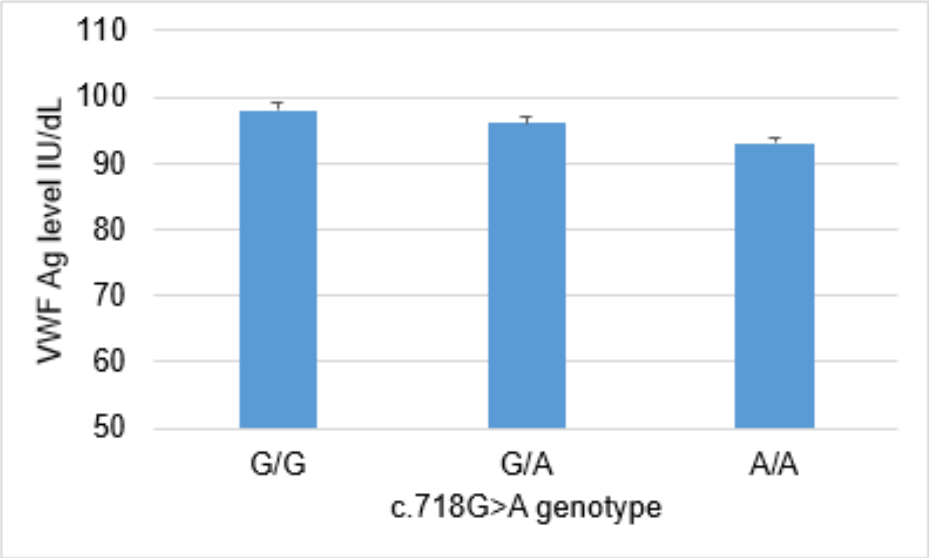


Figure 6-8: Association of SNV c.718G>A with VWF plasma level in all HC (n=921). No significant differences were observed between the three genotypes ($p>0.05$) using the One-way ANOVA (Kruskal-Wallis test). The t-test (Mann-Whitney test) was used to compare between two groups. Bars indicate standard error. Hardy-Weinberg p=0.82.

6.4.1 Investigation of SNVs in exon7 of *CLEC4M*

A previous study had suggested that *CLEC4M* exon 7 encodes a.a involved in ligand binding, suggesting that this exon is of high importance to the function of encoded L-SIGN (Guo*, et al* 2004). Therefore, sequencing of this exon was conducted to identify further variants that might alter binding affinity of L-SIGN, and therefore the clearance of VWF. *CLEC4M* exon 7 was amplified and sequenced in 20 randomly selected HC to look for any variant that might influence binding of L-SIGN. However, no SNV were identified.

6.5 *CLEC4M* VNTR analysis

As mentioned, a previous report suggested that the VNTR in the neck region of L-SIGN was associated with VWF plasma level (Rydz*, et al* 2011). However, this study failed to differentiate the effect of different VNTR sizes on VWF plasma level due to a lack of sufficient individual samples available for study. The decision was therefore taken to investigate the influence of this VNTR further utilising the large EU HC cohort (n=1160) available in Sheffield.

The VNTR was genotyped and to confirm that the genotyping was accurate, two randomly selected samples were amplified and run on an electrophoresis gel (Figure 6-9). The results from the gel matched the results obtained from the genotyping, therefore, the genotyping was presumed to be accurate.

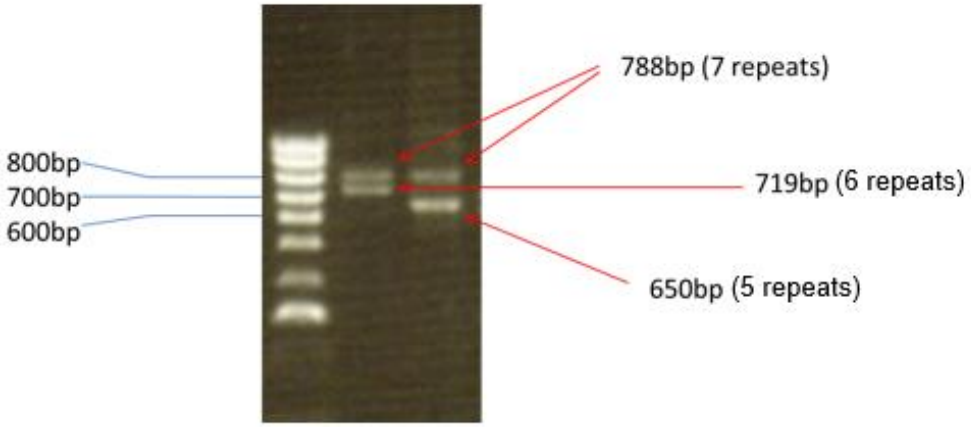


Figure 6-9: The figure shows two HC heterozygous for two different *CLEC4M* VNTR repeat sizes.

6.5.1 VNTR stability

Before investigating the association of the VNTR with VWF plasma levels, the VNTR stability was tested to ensure that the VNTR was inherited between generations without any random addition or removal of one or more VNTR repeats. This was studied by genotyping the VNTR in two families, each containing at least three generations which could be genotyped. The results indicate that the VNTR was stably inherited in these families (Figure 6-10).

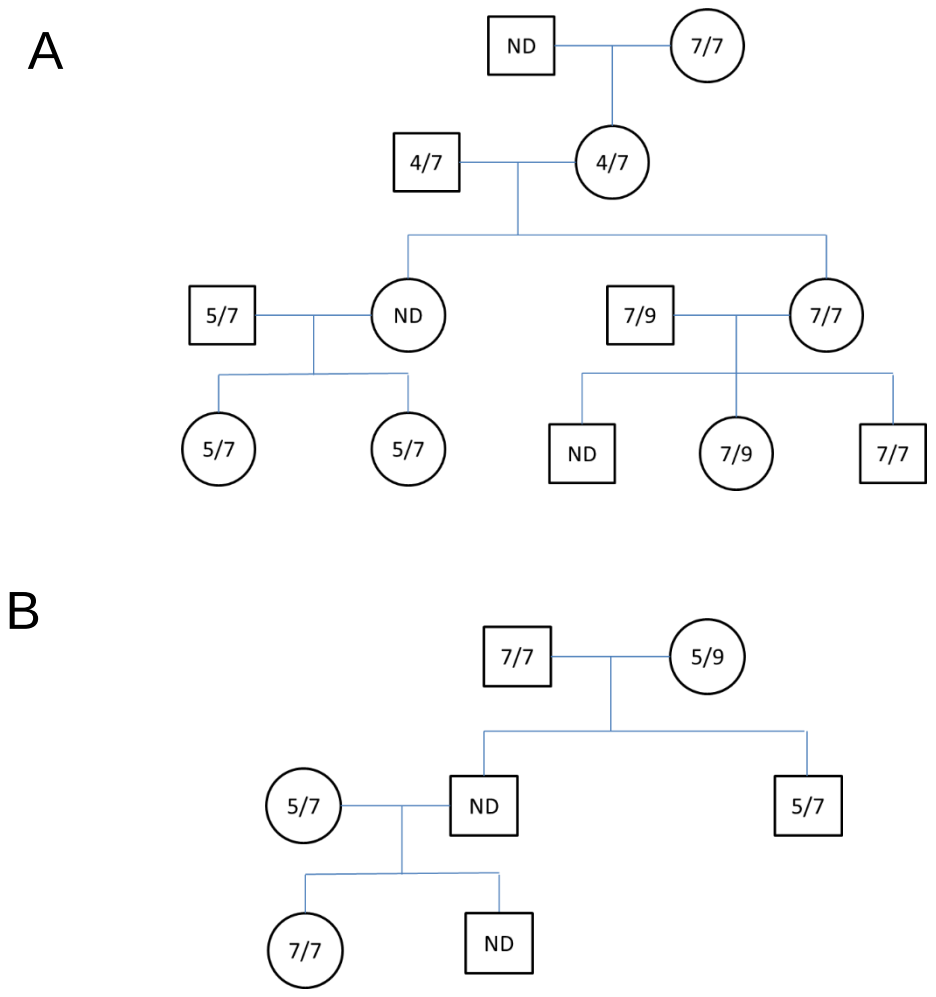


Figure 6-10: VNTR size in the two families (A, P6F16; B, P12F3). ND= DNA not available).

6.5.2 VNTR association with VWF plasma level

The VNTR was successfully genotyped in 921 HC. In addition, genotypic and phenotypic data for 120 Canadian HC was kindly provided by Prof. Lillicrap (Queens University, Kingston, Canada). Before proceeding with analysing the effect of the VNTR from the combined Canadian and EU HC population, the median VWF:Ag levels observed in the two populations were compared to ensure that the results were not affected by any population differences. The results indicated that both populations had comparable VWF:Ag levels (Figure 6-11) (median Canadian VWF:Ag 96 IU/dL; median EU VWF:Ag 98 IU/dL), allowing data from both populations to be combined.

Sizes of VNTR range between 4-9 repeats. Due to relatively small population size available for study, having 21 different genotypes (i.e. six homozygous for a single VNTR size and 15 heterozygous for two different size repeats) would make it difficult to observe any significant association, especially given that some VNTR sizes (such as 4 and 9) were present in a very small number of individuals. Therefore, the decision was to make two groups based on small = (4-6) and large = (7-9) repeats where the population was divided into three main group (small/small, small/large, and large/large) and numbers of individuals were equally distributed between the three groups.

This classification was chosen as 6-7 repeats were the most common repeat size so it was selected as a cut-off point as it would provide balanced distribution of the population samples between different groups (small/small= 239 individuals; small / large= 524 individuals; large/large= 305 individuals). There was a significant association between VNTR size and VWF level ($p=0.038$). Individuals homozygous for the small repeats had a ~10% lower VWF level compared to individuals homozygous for the large repeats (Figure 6-12). Although not significant, the association was similar after taking into account ABO blood group with ~8% lower VWF plasma level for the small/small group when compared to the large/large group for both O (71 IU/dL and 79 IU/dL respectively) and non-O (101 IU/dL and 109 IU/dL respectively) blood group individuals (Figure 6-13). In all instances the genotypes were shown to be in Hardy-Weinberg equilibrium.

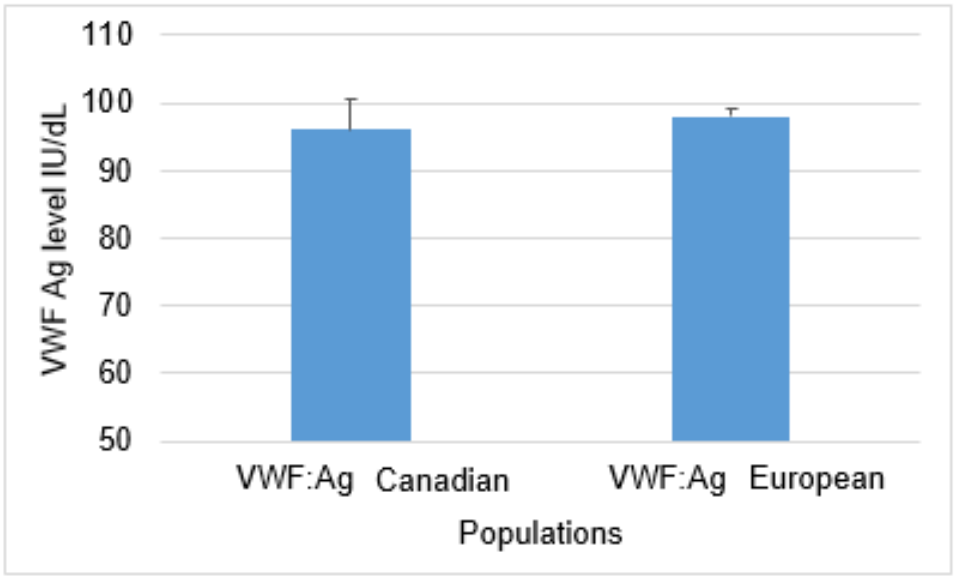


Figure 6-11: Median VWF plasma in European and Canadian HC. Bars indicate standard error.

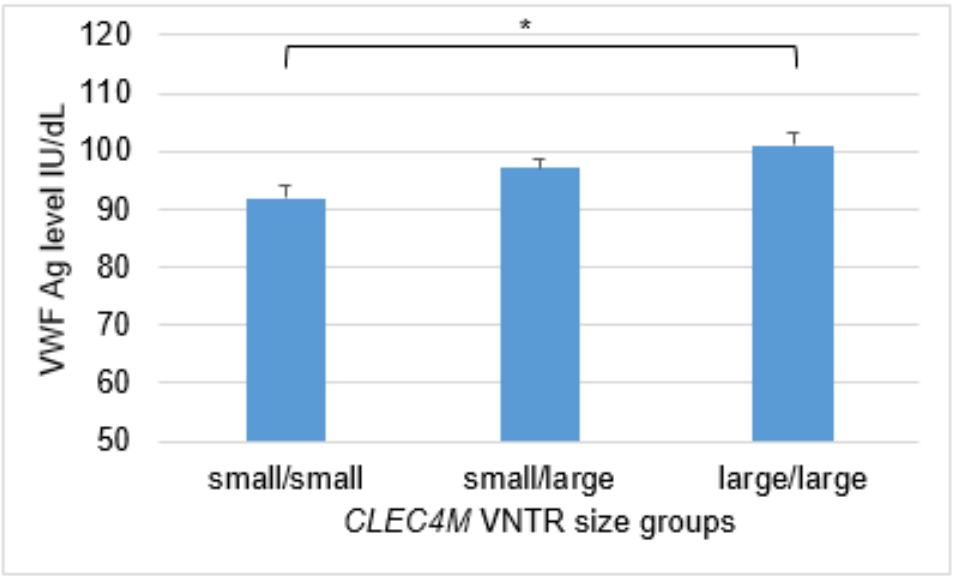


Figure 6-12: Association of *CLEC4M* VNTR size with VWF plasma level. Small, 4-6 VNTR repeats; large, 7-9 repeats. Significant differences were observed between the three alleles (p=0.038) using the One-way ANOVA (Kruskal-Wallis test). The t-test (Mann-Whitney test) was used to compare between two groups; a significant difference was observed between small/small and large/large groups (p=0.012) Bars indicate standard error.

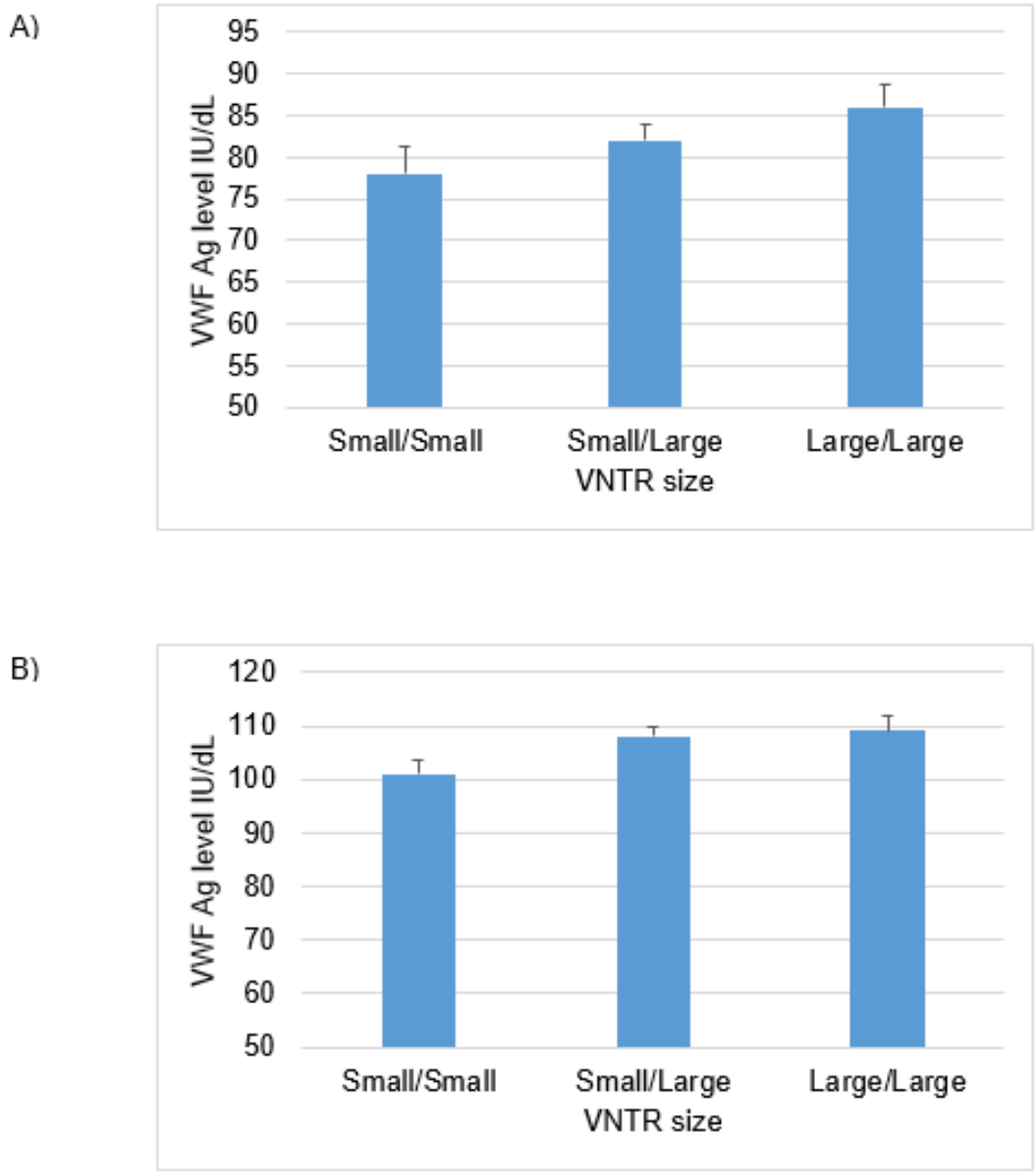


Figure 6-13: The association of different VNTR sizes with VWF plasma level. A) O blood group individuals and **B)** non-O blood group individuals. No significant differences were observed using the One-way ANOVA (Kruskal-Wallis test) to compare the three groups nor the t-test (Mann-Whitney test) was used to compare between two groups. Bars indicate standard error.

6.5.3 Association of homozygous and heterozygous VNTR genotypes with VWF level

It had been reported by a previous study that L-SIGN tetramer with different VNTR sizes (heterozygous VNTR size) are less active than tetramer with a similar VNTR size (homozygous VNTR size) (Bashirova*, et al* 2001). Therefore, VWF levels were also compared between HC who were homozygous and HC who were heterozygous for VNTR genotypes. Analysis of the data obtained in this study showed that homozygous HC had a 5% higher VWF level compared to heterozygous HC, although the results were not significant (Figure 6-14).

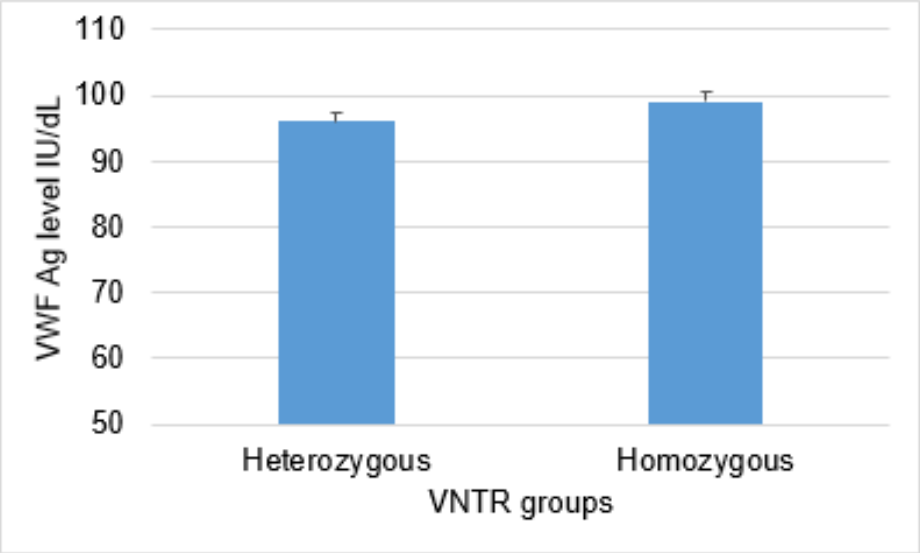


Figure 6-14: Association of homozygous and heterozygous *CLEC4M* VNTR genotypes with VWF plasma level. No significant differences were observed between the two groups ($p>0.05$) using the t-test (Mann-Whitney test). Bars indicate standard error.

6.5.4 Association between *CLEC4M* VNTR and *VWF* SNV c.2385T>C and c.2365A>G with VWF plasma level

The *CLEC4M* VNTR and *VWF* SNV c.2385T>C and c.2365A>G had all been shown to be significantly associated with VWF plasma level. Therefore, further analysis was conducted to assess whether they had an overall additive effect on VWF level (Figure 6-15). The results suggested that these variants did have an additive effect on VWF level when both the non-reference *VWF* SNV alleles and the large VNTR repeats were present in the same individual. Although no mechanism interaction between VWF and CLEC4M variants is expected.

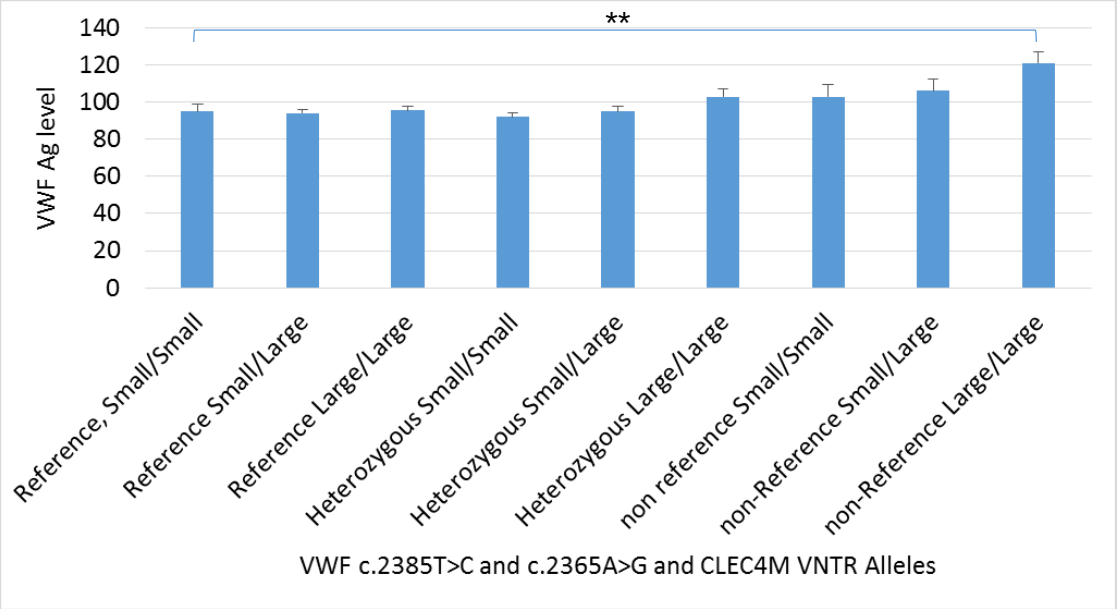


Figure 6-15: The graph shows the effect of different combinations of the *CLEC4M* VNTR and *VWF* SNV (c.2385T>C and c.2365A>G) genotypes on VWF level. VWF Reference in the figure refers to reference allele from both c.2385T>C and c.2365A>G (three individuals were excluded from this analysis due to having different alleles between both SNV), non-reference also refers to non-reference allele for both SNV. For CLEC4M VNTR, small refers to 4-6 repeats while large refers to 7-9 repeats. Significant differences were observed between reference, small/small (95 IU/dL) and non-reference, large/large (121 IU/dL) (p=0.003) using the One-Way-ANOVA (Kruskal-Wallis multiple comparison test). Bars indicate standard error.

6.5.5 Association between *CLEC4M* VNTR and VWF clearance

CLEC4M has been suggested to influence VWF clearance from the circulation (Rydz*, et al* 2013). The rate of VWF clearance can be measured using the VWFpp/VWF:Ag ratio; a higher ratio indicates faster clearance of VWF from the circulation. Therefore, the association of VNTR size with VWFpp/VWF:Ag ratio was also investigated. The results did not show any significant differences between VWFpp/VWF:Ag ratio in the different VNTR groups and thus no significant effect on clearance (Figure 6-16).

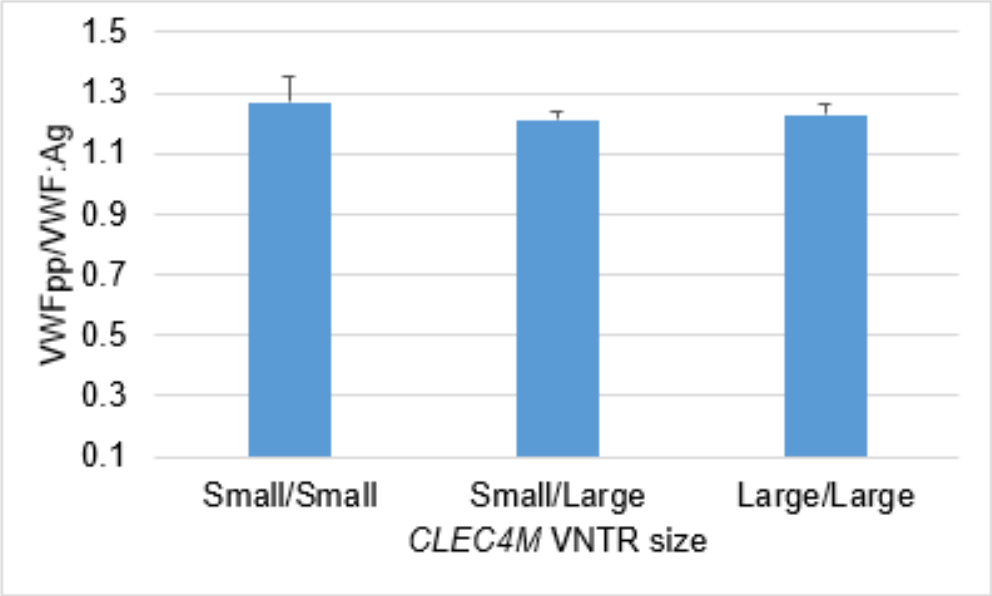


Figure 6-16: Association of *CLEC4M* VNTR on VWFpp/VWF:Ag ratio in ~280HC. No significant differences were observed between the three alleles ($p>0.05$) using the One-Way ANOVA test (Kruskal-Wallis test). The t-test (Mann-Whitney test) also showed no significance comparing between two groups. Bars indicate standard error.

6.6 LD between different *CLEC4M* genotypes

After identifying several variants within *CLEC4M* (other than the CHARGE SNV) with a possible association with VWF plasma level, the LD between these variants and the CHARGE SNV was analysed (Figure 6-17). As mentioned, the data highlighted strong LD between SNV c.718G>A and the CHARGE SNV (c.631+73A>G; 98.5%). However, the CHARGE SNV was also in strong LD with the VNTR when 4-5 repeats were classified as small and 6-9 repeats were classified as large (86.5%; Figure 6-17).

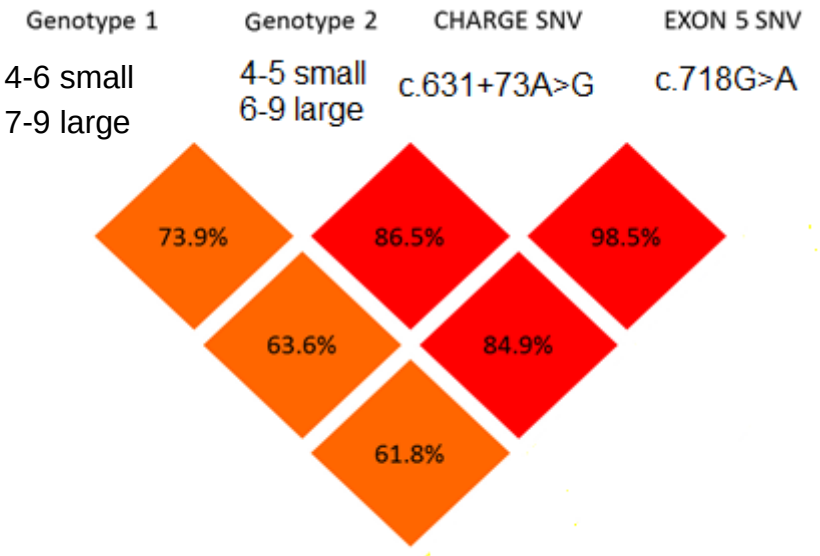


Figure 6-17: Observed LD between different *CLEC4M* genotypes.

Changing the groups of the VNTR from small=4-6 and large=7-9 to small=4-5 and large=6-9 showed a greater LD with the SNV identified by the CHARGE association study. The association between the VNTR (using the new classifications) and VWF plasma level was investigated. This showed a significant VWF increase of ~11% (Figure 6-18).

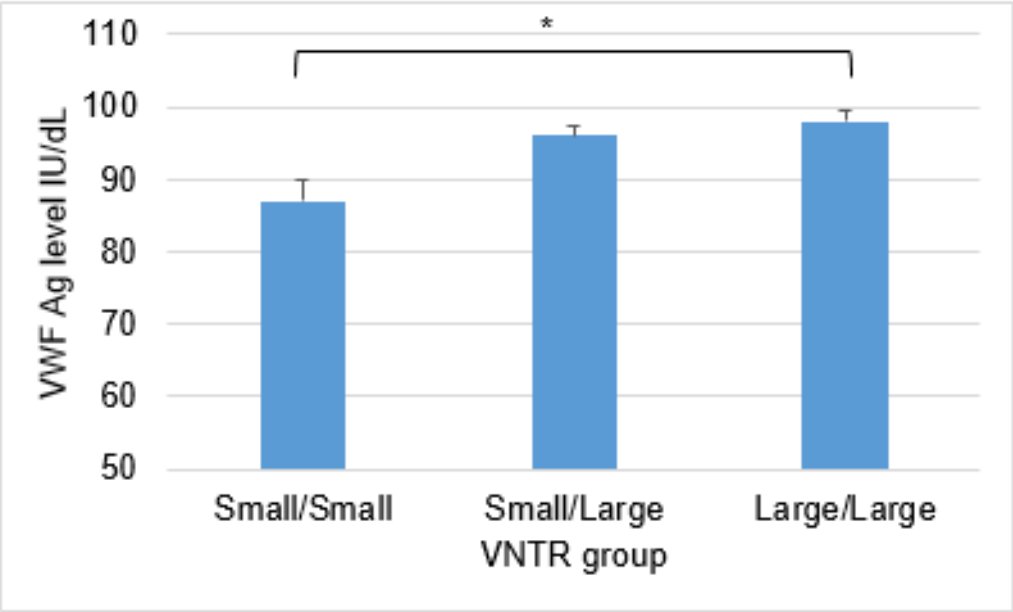


Figure 6-18: Association of *CLEC4M* VNTR size with VWF plasma level. Small, 4-5 VNTR repeats; large, 6-9 repeats (Hardy-Weinberg p=0.23). Significant differences were observed between the three alleles (p=0.031) using the One-Way ANOVA (Kruskal-Wallis test). Significance differences were also observed between Small/Small and Large/Large repeats using the t-test (Mann-Whitney test) (p=0.011). Bars indicate standard error.

6.7 *CLEC4M* expression

To assess the possible effect of *CLEC4M* on VWF level, investigation of the expression of *CLEC4M* in HUVECs was conducted. HUVECs DNA and mRNA from two individuals were amplified using two sets of primers due to the presence of repetitive sequences and the presence of several splice variants. All the primers were designed to be in the middle of the exon to amplify RNA as well as DNA.

CLEC4M DNA was amplified using both primers sets, while *CLEC4M* mRNA was not detected. The results suggest the lack of expression of *CLEC4M* in these cells (Figure 6-19). Therefore, *CLEC4M* expression appears to be limited to liver endothelial cells where VWF clearance has been reported (Denis*, et al* 2008).

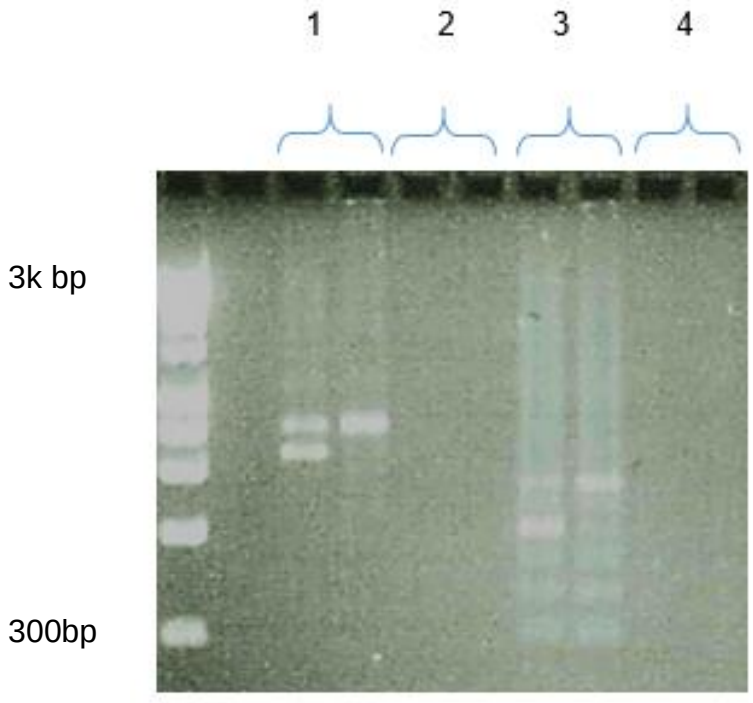


Figure 6-19: Investigation of *CLEC4M* expression in HUVECs from 2 individuals using 2 primers sets. 1) DNA with primer set 1; 2) mRNA from primer set 1; 3) DNA with primer set 2 (non-specific bands are seen due to no specificity for the primers used); 4) mRNA with primer set 2.

6.8 *FUT3* and *CLEC4M*

From the previous findings, *FUT3* and *CLEC4M* are significantly associated with VWF plasma level. The two genes are present within 2Mb (Figure 6-20) on chromosome 19. Therefore, the LD between the two genes was investigated using HapMap data, however, no LD data was available (Figure 6-21). Our data for the *CLEC4M* VNTR and *FUT3* variant was used to investigate the LD between them, however, there was no LD identified between the two (r^2= 0.032).

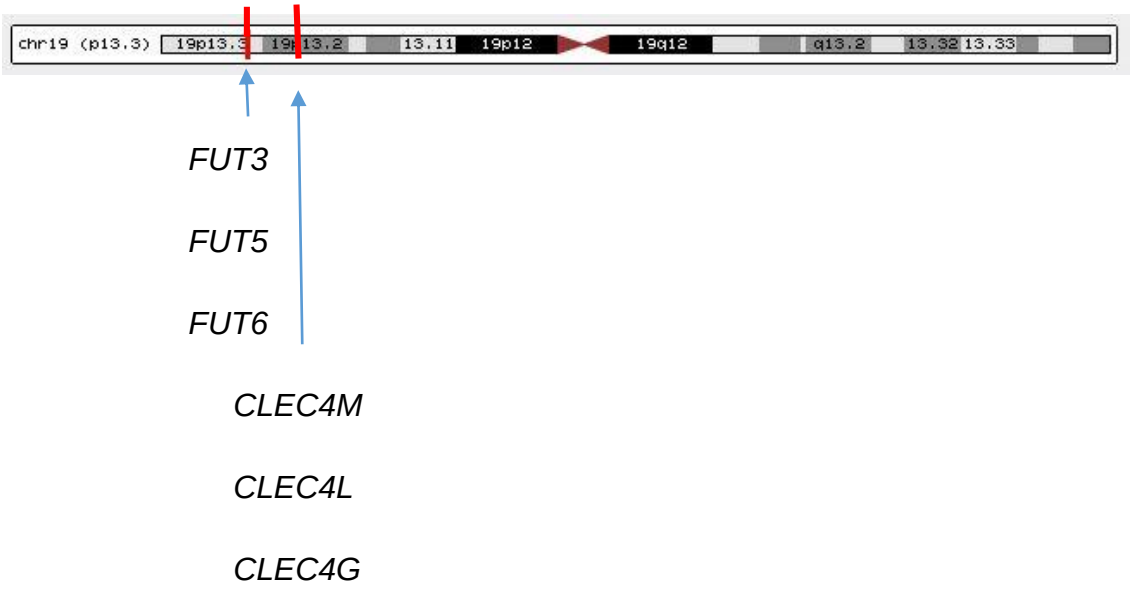


Figure 6-20: Location of *FUT3* and *CLEC4M* on chromosome 19.

Figure 6-21: LD between *CLEC4M* SNV and *FUT3* c.202T>C variant (shown in red oval). Completely red squares represent 100% LD; completely grey squares represent missing data, light red squares refer to weaker LD while white squares refer to lack of LD)

6.9 Discussion

The aim of this chapter was to investigate the effect of *CLEC4M* on VWF levels. A previous association study identified a variant (c.631+73A>G) in *CLEC4M* with significant association with VWF plasma level (Smith*, et al* 2010). It was suggested that L-SIGN (encoded by *CLEC4M*) contributes to the clearance of VWF from the circulation (Rydz*, et al* 2013). Therefore, this study investigated the association of *CLEC4M* with VWF level further. This investigation showed a significant association between the *CLEC4M* VNTR with VWF plasma level.

Using the Gene Cards web tool, the possible proteins potentially involved in the *CLEC4M* pathway, i.e. being co-expressed or having a physical interaction with *CLEC4M* were investigated to try and identify any known proteins that were also previously reported to interact with VWF which could possibly provide an indication of a possible mechanism of interaction between L-SIGN and VWF. However, the results did not suggest any known mechanism of interaction (Figure 6-22). Further investigation using the Genemania web tool gave similar results (data not shown).

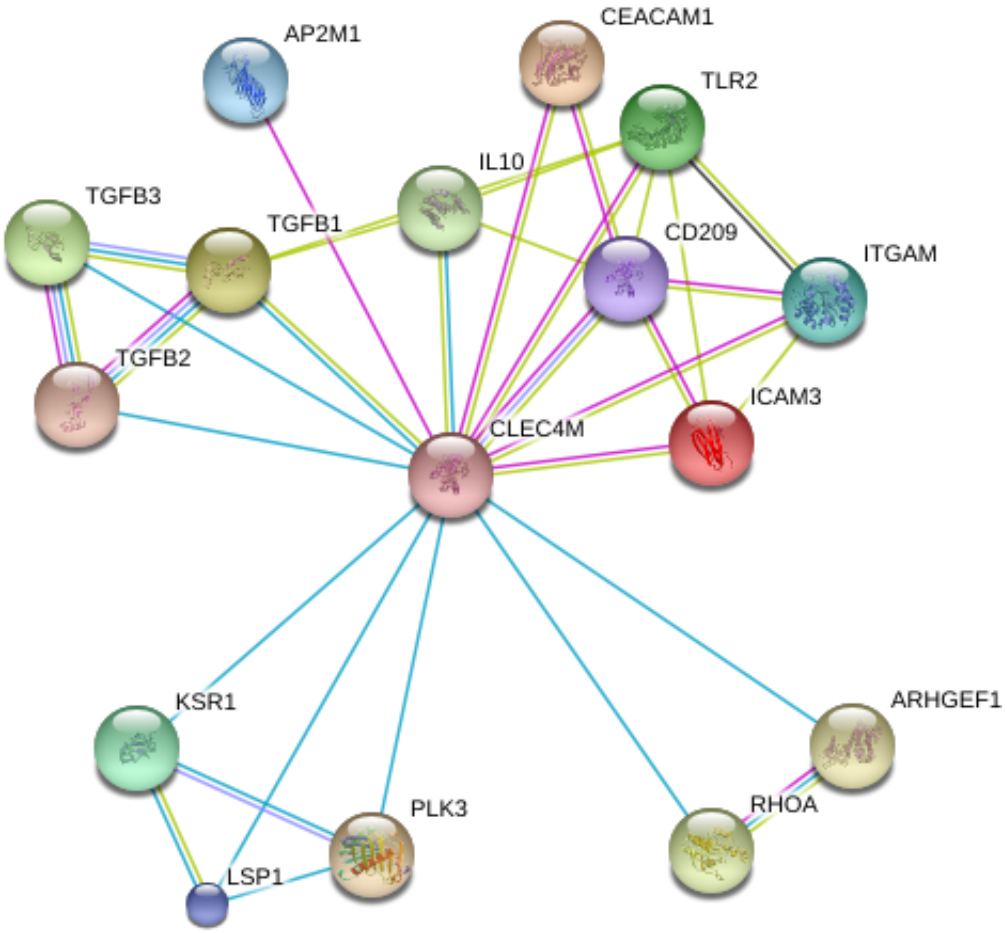


Figure 6-22: Proteins potentially involved in the *CLEC4M* pathway. Proteins that physically interact with L-SIGN (purple connections), involved in a similar pathway (blue connections) described in the same reference publication (yellow connections) sharing homology (light blue connections) and neighbouring proteins (green connections).

Three main regions of *CLEC4M* were highlighted to be of importance to L-SIGN function and therefore potentially influencing VWF level. Firstly, exon 7 which encodes a.a directly involved in L-SIGN ligand binding (Guo*, et al* 2004). Secondly, LD investigation identified two SNV (c.585G>A in exon 4 and c.718G>A in exon 5) with strong LD to the CHARGE SNV. Thirdly, the VNTR in exon 4 that had been previously suggested to influence the binding affinity of L-SIGN (Feinberg*, et al* 2005, Rydz*, et al* 2011).

Exon 7 was sequenced in 20 random HC, but no SNV was found in any of them (completely WT), suggesting that this exon in the European population has no variation (which was confirmed using HapMap data). However, populations from other ethnic backgrounds might still have SNV of functional importance. Nevertheless, this region was excluded from any further investigations.

c.585G>A was not predicted to have any effect on L-SIGN using four different protein prediction tools and was excluded from further investigation. c.718G>A was predicted to have an effect using two out of four tools and was investigated further. While the results showed that individuals with the non-reference allele had ~5% reduction in VWF level compared to individuals with the reference allele this association was not significant.

The CHARGE study identified c.631+73A>G to have the weakest significant association with VWF plasma level (compared to other identified SNV in different genes) in a much larger population and our data showed that c.718G>A is not in perfect LD with c.631+73A>G. In addition, variants with a known association to VWF level (such as ABO blood group) could not be taken into account in this investigation due to the small population size. It is therefore not unexpected to see a lack of significant association in our population which does not mean that the association is absent.

Genotyping the VNTR in the EU HC combined with previous genotyping data from Canadian HC showed significant association between the *CLEC4M* VNTR and VWF plasma level. The suggestion is that the association is weak given that data was required from two study populations, which is similar to

what was observed with the CHARGE study. In addition, this supports the possibility that a significant association of c.718G>A with VWF level could be seen if investigated using a larger population sample (no genotypic data was available for this SNV from the Canadian population).

A previous study had suggested that tetramer L-SIGN with each subunit having the same size (i.e. *CLEC4M* homozygous for one VNTR repeat) were more stable and more efficient at binding targets compared to any heterozygous tetramers (Bashirova*, et al* 2001). This theory was investigated in our study, with the expectation that higher VWF levels would be observed in individuals heterozygous for *CLEC4M* VNTR size due to lowered VWF clearance efficiency. However, the opposite trend (although not significant) was observed. This does not completely abolish the theory as suggested by the previous study because in our study different VNTR sizes were combined together which would have an impact on the observed association as comparing each group alone would give more accurate indication of the association. An ideal study would be to compare homozygous and heterozygous repeats of each size together (for example comparing 4/4, 4/5 and 5/5 together, and the same for all sizes) which was not done in this study due to the small population sample that made this impractical.

Due to the similarity of the observed association between the *CLEC4M* VNTR and *VWF* variants c.2385T>C / c.2365A>G with VWF levels, the decision was taken to investigate whether the associations were linked in some way or whether they had completely separate effects. The results suggest that these variants are independently influencing VWF plasma level and combining them demonstrates an additive effect on VWF level. This in turn suggests that *VWF* variants c.2385T>C / c.2365A>G and *CLEC4M* influence VWF level independently and the presence of both variants combined lead to a greater influence on VWF level.

Other C-type lectin family members including *CLEC4L* and *CLEC4G* were found to be present within close proximity to *CLEC4M* and share some homology with *CLEC4M*. In addition, a previous study showed that *CLEC4L,* expressing DC-SIGN interacts with several antigens that form part of the

glycosylated glycans present on VWF such as ABH antigens (Guo*, et al* 2004). HapMap data suggested no LD between *CLEC4M* with either *CLEC4L* or *CLEC4G*. However, this does not deny any possible association between L-SIGN and VWF, but could instead, suggest that if there is any association, it will be independent of the L-SIGN association. Due to time limitations, no further work was conducted to study any possible association between DC-SIGN (encoded by *CLEC4L*) and VWF levels. In addition, the lack of interaction between *CLEC4M* and some antigens does not implicate lack of direct interaction with VWF as the interaction could be with glycans not included in this study or with VWF protein itself (Guo*, et al* 2004).

Following that, the association between *FUT3* c.202T>C variant and *CLEC4M* VNTR was analysed looking for possible LD between the two variants. However, no LD was observed suggesting that these two variants influence VWF plasma level through independent pathways.

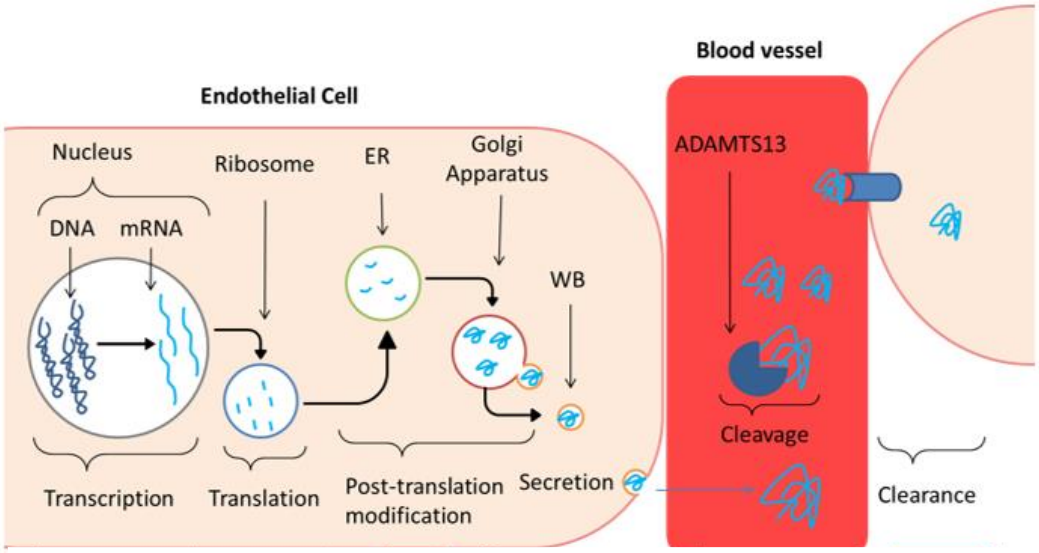
In summary, this work provided evidence of association between *CLEC4M* VNTR with VWF plasma level. It also suggested a SNV (c.718G>A) with a potential effect on VWF plasma level that requires investigation of additional samples to achieve significance. A previous study showed no interaction between L-SIGN and some antigens glycosylated on VWF but this does not deny the association (Guo*, et al* 2004). Recent work has provided more evidence supporting the interaction and association between *CLEC4M* and VWF level (Rydz*, et al* 2011, Rydz*, et al* 2013). Finally, these findings support the hypothesis that the VNTR in *CLEC4M* is in strong LD with the SNV identified by the CHARGE study suggesting an association with VWF level through alteration in the neck size.

7 Chapter seven: General discussion

7.1 Discussion

VWF is large plasma glycoprotein that plays an important role in blood coagulation by enhancing platelet adhesion and aggregation to sites of injury and carrying FVIII and protecting it from degradation. VWF level varies between 50-200 IU/dL in the normal population (Gill*, et al* 1987). Several factors are known to influence VWF plasma level including both genetic factors, such as ABO blood group, and environmental factors such as age and stress. In addition to the variation of VWF plasma level in the normal population, VWF plasma level also varies between VWD patients with the same causative mutations. In addition, some individuals exhibiting a VWD phenotype were found to not have any causative *VWF* mutations, suggesting that an additive effect of several variants influencing VWF level may lead to a VWD-like phenotype. The alternate allele of these variants can also be considered risk factors for cardiovascular disease, i.e. increasing VWF levels.

Given that VWF is known to undergo extensive post-translational modifications and interacts with a large number of different protein ligands during its circulation (Lenting*, et al* 2015), it is expected that a large number of genetic factors might influence VWF level (Figure 7-1). Although the effect of several genetic factors influencing VWF levels has been established, others are suggested to influence VWF level but need further investigation. This study initially aimed to identify additional genetic loci and variants associated with VWF plasma level that are responsible for VWF level variations reported in the normal population and then to investigate their mechanism of action.



VWF	*VWF*	*ABO, FUT3, FUT1, FUT2*	*AFTPH, SYNRG, CLTCL1, RAB27a, STXBP5, STX2*	*ADAMTS13*	*CLEC4M, Siglec-5, STAB2*

Figure 7-1: Life cycle of VWF. Genes encoding proteins involved at different stages of the VWF life cycle are highlighted.

Several genome wide association studies identified genetic loci both within and other than *VWF* to be significantly associated with VWF plasma levels (Souto*, et al* 2003, Smith*, et al* 2010, Zabaneh*, et al* 2011). These studies provided evidence that genetic loci other than *VWF* are associated with VWF plasma level. In addition, they provided evidence that variants present in the normal population, previously referred to as neutral polymorphisms, also have an effect on VWF levels. However, none of these reports or any other study investigated the mechanism by which a "polymorphism" could be influencing VWF plasma level.

This work also aimed to complement previous studies which were conducted to identify mutations in *VWF* associated with VWD1 in the European and Canadian populations (Goodeve*, et al* 2007, James*, et al* 2007). These studies failed to identify the causative mutations in a subset of patients diagnosed with VWD1 suggesting either that other genetic loci outside of *VWF* could be associated with VWF level or that possibly a combined effect of two or more SNVs, not disease causing on their own, led to a VWD1-like phenotype. In addition, a better understanding of other genetic factors that influence VWF level will lead to better understanding of the variation of VWF level in patients with the same causative mutations.

There have been numerous published works highlighting an association of several *VWF* variants with VWF level either as part of genome wide association studies or *VWF* specific studies (Souto*, et al* 2003, Smith*, et al* 2010, Campos*, et al* 2011, Zabaneh*, et al* 2011, Zhou*, et al* 2014). This study confirmed the association of previously described *VWF* variants with VWF level including c.2385T>C and c.2365A>G. In addition, it also identified variants that have not been suggested to have an association with VWF levels such as c.7888-3501A>G, c.7888-2310T>C and c.-1342C>G, which due to time limitations were not studied in detail as part of this work. Most of the *VWF* variants identified in this study were intronic or silent exonic variants supporting previous findings regarding the importance of these type of variants which were reported as the majority of variants influencing VWF level (Campos*, et al* 2011). However, as of yet, there is no published work

investigating the mechanism by which these variants influence VWF levels. Therefore, the next step was to investigate their mechanism of action.

Both c.2385T>C and c.2365A>G were found to be associated with ~9% increase in VWF level compared to the reference allele in this study as well as in an earlier study (Zabaneh*, et al* 2011). *In vitro* investigation showed that they are also associated with a significant increase in protein expression, mRNA expression and mRNA half-life. However, they were not found to have any effect on splicing nor were they predicted to have any effect on mRNA secondary structure or on miRNA binding motifs, which are the three main factors influencing mRNA half-life (Duan*, et al* 2013). However, the same study suggested that a change of nucleotide from A or T to C or G would increase mRNA half-life by an unknown mechanism which could be the reason for the observed increase on mRNA half-life observed in this study (Duan*, et al* 2013).

VWF c.2771G>A and c.4146G>C variants were identified by this research group and by others to be significantly associated with VWF plasma levels (Goodeve*, et al* 2007, James*, et al* 2007, Corrales*, et al* 2012). Therefore, this study aimed to investigate their mechanism of action. Previous studies had ruled out an effect of c.2771G>A on VWF multimerisation, storage or FVIII binding but none looked at its effect on mRNA level (Hilbert*, et al* 2003, Berber*, et al* 2009, Hickson*, et al* 2010). Prediction of the effect of c.2771G>A and c.4146G>C on ESE motifs showed a possible effect and *in vitro* investigation confirmed an alteration of ESE motifs caused by these variants. The effect of both SNV in mRNA and mRNA half-life showed a reduction suggesting that the reduction in mRNA half-life is due to a defect in splicing through disruption of ESE sites supporting previous findings for c.2771G>A (Berber*, et al* 2009). The effect of intronic *VWF* variant c.1534-3C>A was previously reported to causes a defect in splicing and therefore VWD1 which would support our finding that defect in splicing could lead to VWD1 (Gallinaro*, et al* 2006).

Previous studies suggested an association between Lewis blood group antigens and VWF level (Orstavik*, et al* 1989, Green*, et al* 1995, Hickson*, et al* 2009) and provided evidence against what was suggested by another study that did not see these associations. However, this study used a much smaller

sample size which could be the reason for their observed lack of association (O'Donnell*, et al* 2002). The conversion of the precursor glycan chain to Lewis antigen is performed by α-3-fucosyltrasferase (encoded by either *FUT3* or *FUT6*). The c.202T>C variant in *FUT3* was reported to be specifically associated with VWF plasma level (Hickson*, et al* 2009). However, *FUT3* expression had only been reported in exocrine secretions while *FUT6* (which shares 84% homology and is located within 40kb of *FUT3)* is expressed in plasma (Mollicone*, et al* 1994, Costache*, et al* 1997).

Investigation of LD between *FUT3* c.202T>C and SNVs in *FUT6* and additionally *FUT5* (also highly homologous and present in close proximity to *FUT3*), revealed that no other SNVs were in strong LD. Therefore, the c.202T>C variant in *FUT3* seems to be entirely responsible for the observed association with VWF level. Although *FUT3* has not been reported to be expressed in plasma, this could be explained by the high degree of homology between *FUT3* and *FUT6*; *FUT6* expression could obscure plasma expression or detection of *FUT3*. Even if *FUT3* is not expressed in plasma, it might still be associated with VWF level through an indirect effect. Further work is therefore required to investigate the mechanism by which *FUT3* expression influences VWF plasma level.

This study also investigated the proposed association of *CLEC4M* with VWF plasma level. A variant (c.631+73A>G) in *CLEC4M* was reported to be significantly associated with VWF plasma level (Smith*, et al* 2010). Analysing SNVs in strong LD with this SNV identified two other SNVs in *CLEC4M* (c.585G>A, and c.718G>A). Only c.718G>A was predicted to affect L-SIGN structure, further analysis of SNV c.718G>A revealed that it was present in exon 5 (encoding the outer surface of the extracellular domain) which suggested that the SNV might influence ligand binding. Analysis of this SNV showed a trend towards a ~5% variation in VWF level with the non-reference allele associated with higher levels, thus an increase in L-SIGN affinity towards possible binding and clearance of VWF. However, the results did not achieve significance suggesting that data from additional samples was needed in order to achieve significance.

Several studies have also reported an association between a VNTR within the neck region of L-SIGN and receptor affinity (Soilleux*, et al* 2000, Guo*, et al* 2004, Chan*, et al* 2006), and a recent study reported an association between this VNTR and VWF level (Rydz*, et al* 2011). Analysis of the VNTR performed in this study suggested that large alleles of the *CLEC4M* VNTR (7-9 repeats) were associated with an ~10% increase in VWF plasma level compared to small VNTR alleles (4-6 repeats). The observed trend was visible both before and after taking into account ABO blood group, but additional samples need to be added to the study to achieve significance when taking into account ABO blood group.

L-SIGN forms tetramers on the cell surface. A previous study reported that L-SIGN tetramers with the same VNTR size had a higher binding affinity than tetramers of different VNTR sizes (Soilleux*, et al* 2000). However, investigation of VWF level between individuals homozygous and heterozygous for the VNTR showed that homozygous individuals had ~5% higher VWF levels which is not in line with a previous report (Bashirova*, et al* 2001), which could be due to dividing all similar genotypes into two groups (i.e. individuals homozygous for any VNTR size were one group and heterozygous individuals was the other group). Therefore the differences were not clearly shown. The small sample size did not allow individuals to be divided into sub-groups.

The LD between the *CLEC4M* CHARGE SNV (c.631+73A>G), exon 5 SNV (c.718G>A) and the VNTR was also investigated in our population. The results showed that all were in strong LD (Figure 6-17). This might suggest that these variants are having an additive effect (because the variants have a similar effect on VWF level and are in strong LD), or potentially only one variant has an effect but due to the strong LD, the other variants showed a similar association. Further functional work is still required to test whether these variants have a separate or combined effect.

Both *FUT3* and *CLEC4M* are suggested to be associated with VWF clearance (Hickson*, et al* 2009, Rydz*, et al* 2013). Little is currently known about the VWF clearance mechanism. Lewis antigens were predicted to be associated with glycosylation and clearance in the same way as ABH antigens. However,

recent study showed evidence that Lewis antigens are not glycosylated on VWF (Canis*, et al* 2012), nevertheless, *FUT3* might still be associated with VWF level either through glycosylation of other proteins that interact with VWF (such as ADAMTS13 which is reported to be glycosylated) or it directly interacts with VWF in an unknown way. It has been reported that the liver (which is reported to express L-SIGN) plays an important role in VWF clearance (van Schooten*, et al* 2008). Macrophages are also involved in VWF clearance (van Schooten*, et al* 2008) and are reported to be recruited by L-SIGN to clear glycosylated products (Guo*, et al* 2004). L-SIGN clears glycosylated and fucosylated products such as viruses, bacteria and ABH and Lewis antigens (Guo*, et al* 2004, Zandberg*, et al* 2012) and is predicted to clear VWF in a similar way.

7.2 Study limitations

The current study, like any other study, has some limitations. As the study looked at three main genes, the limitations of looking at each gene will be discussed separately. First, the analysis of *VWF* only covered a small number of SNV (111) within a gene that is known to be highly polymorphic and this means that a large number of variants were overlooked, which does not mean they are of less importance. However, due to limited time and costs, the analysis was only conducted on the available data. In addition, several SNV identified in this study that were found to be significantly associated with VWF level including c.7888-3501A>G, c.7888-2310T>C and c.-1342C>G were excluded from further investigation due to time limitations. Furthermore, all of the mRNA splicing work was done *in vitro*, which may not entirely reflect human mRNA, but due to the lack of access to individual RNA samples carrying these variants, such investigation was not performed.

The main weakness when looking at *FUT3* and *CLEC4M* work was that some areas (i.e. introns, UTRs and the promoter region) within *FUT3*, *FUT5*, *FUT6* and *CLEC4M* were not analysed and variants may therefore have been missed that could also possibly be associated with VWF plasma level. In addition, only variants found to be in strong LD with *FUT3* c.202T>C and *CLEC4M* c.631+73A>G were considered for further investigation, therefore,

other variants with an independent effect on VWF level may have been missed.

7.3 Future work

The current association analysis was performed on HC. Similar analysis could be performed on VWD1 patients, especially those that have similar causative mutations but show variation in VWF level within the same family, in addition to patients who were not identified to have *VWF* mutation(s) or the identified mutations did not fully explain their phenotype. This would be done by investigating whether these patients carry any of the previously reported variants in *VWF*, *FUT3* and/or *CLEC4M* and determining whether these variants explain the variation in VWF level and VWD severity. This could indicate whether these variants potentially contribute to the severity of VWD.

c.2385T>C and c.2365A>G are both present within the *VWF* D3 domain that is known to be associated with FVIII binding and VWF multimerisation (Goodeve*, et al* 2007). Both SNV were found to be significantly associated with FVIII level in the previous CHARGE study (Smith*, et al* 2010). Therefore, analysis of the effect of these two SNV on FVIII binding would be undertaken. In addition, multimer analysis would be performed to test the effect of both SNV on VWF multimerisation.

The effect of *VWF* c.7888-3501A>G and c.7888-2310T>C in intron 47 on VWF plasma level still needs to be studied further. From the current *in silico* analysis, c.7888-3501A>G appears to be directly associated with *VWF* splicing by introducing a new donor site (predicted by 4 out of 5 web tools). Therefore, ESE and acceptor / donor motif splicing assays would be used to test the effect of this variant on ESE motifs and splicing (Bonnet*, et al* 2008). c.7888-2310T>C did not seem to be directly influencing VWF level as it was not predicted to have any major effect on ESE motifs, splicing or transcription factor binding. However, surrounding SNV showed a possible effect. The first step in investigating this association further would be to amplify and sequence 1kb of DNA each side of the c.7888-2310T>C variant and look for other SNV

in strong LD with it, then those SNV would be tested *in silico* to predict any potential effect which would then be investigated further.

VWF c.2771G>A and c.4146G>C were found to have an effect on ESE motifs, however, it would be interesting to also test their effect on acceptor / donor motif splicing as it would provide better indication of their effect on splicing as an effect on ESE alone can less likely be of major impact, however, subsequent effect on splicing can be disease causing (Gallinaro*, et al* 2006). This experiment was not performed due to time limitations. In addition, testing the effect of c.2771G>A on splicing using mRNA extracted from blood, which became available at the time of writing this thesis, would be performed which will provide a more accurate indication of the effect of this variant *in viv*o.

Investigation of the association and interaction between L-SIGN and VWF was not carried out during this project due to it being conducted by a group led by Prof. Lillicrap (Queen's University, Kingston, ON, Canada) who work in collaboration with our research group. The results of the association was published earlier, thus, no further investigation was conducted by our research group (Rydz*, et al* 2013). What would have been done was to study the mechanism by which L-SIGN is influencing VWF plasma level. This could be done using plasmids with *CLEC4M* cDNA inserted for different VNTR sizes. Additionally, the effect of c.718G>A (both separately and in combination with different VNTR sizes) would be examined. Following that, different plasmids would have been transfected into HEK293T cells and VWF internalisation would have been compared. This would be done using recombinant VWF (Humate-P) and measured using flow cytometry (looking for changes in the extracellular and intracellular intensity which would be indicate internalisation). An empty plasmid (the same plasmid but without *CLEC4M*) would be used as a negative control to take into account any other receptors that might internalise VWF. Not adding recombinant VWF would also be used as a negative control, to check that what increases the intracellular intensity is the recombinant VWF used.

The association of *FUT3* with VWF level would also be investigated first by examining intronic and promoter SNV in *FUT3*, *FUT5* and *FUT6* to make sure

that the c.202T>C variant was responsible for the observed association with VWF level. Second, co-transfecting plasmids carrying WT of both *VWF* and *FUT3* cDNA and comparing the secreted VWF level with that obtained using plasmid containing WT *VWF* with *FUT3* plasmid carrying c.202T>C variant. In addition, the secreted VWF could be used to test for effect on clearance using other cells expressing *CLEC4M*. This would give an indication of the effect of *FUT3* c.202T>C on VWF expression and clearance.

Furthermore, variants in *FUT5* and *FUT6* were only analysed by looking for SNVs in strong LD with c.202T>C, however some variants might have an independent effect on VWF level. Therefore, *in silico* analysis would be conducted on other variants found in *FUT5* and *FUT6*, then SNV found to have potential effect would be genotyped in more individuals and any significant effect on VWF level examined further.

7.4 Concluding remarks

SNV have historically been believed to be benign, however, there is growing evidence that several SNV are important for gene function or expression. In this study, several SNV were found to be associated with VWF plasma level within *VWF* and other genetic loci including *CLEC4M* and *FUT3*. These findings provide a better understanding of the variation in VWF level in the normal population as well as variation in the bleeding severity observed in VWD patients with the same causative mutation. In addition, although these SNV are not disease causing, they may be potential risk factors for bleeding or for cardiovascular disease and an additive effect can increase the risk further as suggested by this study. This study reported the first case of SNV associated with VWF plasma level through interfering with mRNA half-life. This study has also provided evidence, for the first time, of mutations affecting ESE contributing to VWD1. In conclusion, these findings support the hypothesis that SNV influence VWF level and that genetic loci other than *VWF* influence VWF plasma level.

8 References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. & Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat Methods,* **7,** 248-249.

Barreiro, L.B., Patin, E., Neyrolles, O., Cann, H.M., Gicquel, B. & Quintana-Murci, L. (2005) The heritage of pathogen pressures and ancient demography in the human innate-immunity CD209/CD209L region. *Am J Hum Genet,* **77,** 869-886.

Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics,* **21,** 263-265.

Baruch, D., Denis, C., Marteaux, C., Schoevaert, D., Coulombel, L. & Meyer, D. (1991) Role of von Willebrand factor associated to extracellular matrices in platelet adhesion. *Blood,* **77,** 519-527.

Bashirova, A.A., Geijtenbeek, T.B., van Duijnhoven, G.C., van Vliet, S.J., Eilering, J.B., Martin, M.P., Wu, L., Martin, T.D., Viebig, N., Knolle, P.A., KewalRamani, V.N., van Kooyk, Y. & Carrington, M. (2001) A dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)-related protein is highly expressed on human liver sinusoidal endothelial cells and promotes HIV-1 infection. *J Exp Med,* **193,** 671-678.

Berber, E., James, P.D., Hough, C. & Lillicrap, D. (2009) An assessment of the pathogenic significance of the R924Q von Willebrand factor substitution. *J Thromb Haemost,* **7,** 1672-1679.

Bonfield, J.K., Rada, C. & Staden, R. (1998) Automated detection of point mutations using fluorescent sequence trace subtraction. *Nucleic Acids Res,* **26,** 3404-3409.

Bonnet, C., Krieger, S., Vezain, M., Rousselin, A., Tournier, I., Martins, A., Berthet, P., Chevrier, A., Dugast, C., Layet, V., Rossi, A., Lidereau, R., Frebourg, T., Hardouin, A. & Tosi, M. (2008) Screening BRCA1 and BRCA2 unclassified variants for splicing mutations using reverse transcription PCR on patient RNA and an ex vivo assay based on a splicing reporter minigene. *J Med Genet,* **45,** 438-446.

Bowen, D.J. & Collins, P.W. (2004) An amino acid polymorphism in von Willebrand factor correlates with increased susceptibility to proteolysis by ADAMTS13. *Blood,* **103,** 941-947.

Brinkman-Van der Linden, E.C., Mollicone, R., Oriol, R., Larson, G., Van den Eijnden, D.H. & Van Dijk, W. (1996) A missense mutation in the FUT6 gene results in total absence of alpha3-fucosylation of human alpha1-acid glycoprotein. *J Biol Chem,* **271,** 14492-14495.

Brunak, S., Engelbrecht, J. & Knudsen, S. (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol,* **220,** 49-65.

Campos, M., Sun, W., Yu, F., Barbalic, M., Tang, W., Chambless, L.E., Wu, K.K., Ballantyne, C., Folsom, A.R., Boerwinkle, E. & Dong, J.-F. (2011) Genetic determinants of plasma von Willebrand factor antigen levels: a target gene SNP and haplotype analysis of ARIC cohort. *Blood,* **117,** 5224-5230.

Canis, K., McKinnon, T.A., Nowak, A., Haslam, S.M., Panico, M., Morris, H.R., Laffan, M.A. & Dell, A. (2012) Mapping the N-glycome of human von Willebrand factor. *Biochem J,* **447,** 217-228.

Canis, K., McKinnon, T.A.J., Nowak, A., Panico, M., Morris, H.R., Laffan, M. & Dell, A. (2010) The plasma von Willebrand factor O-glycome comprises a surprising variety of structures including ABH antigens and disialosyl motifs. *J Thromb Haemost,* **8,** 137-145.

Carew, J.A., Browning, P.J. & Lynch, D.C. (1990) Sulfation of von Willebrand factor. *Blood,* **76,** 2530-2539.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. & Krainer, A.R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res,* **31,** 3568-3571.

Castaman, G., Tosetto, A., Eikenboom, J.C. & Rodeghiero, F. (2010) Blood group significantly influences von Willebrand factor increase and half-life after desmopressin in von Willebrand disease Vicenza. *J Thromb Haemost,* **8,** 2078-2080.

Chakrabarti, R., Brozovic, M., North, W.R., Stirling, Y. & Meade, T.W. (1975) Effects of age on fibrinolytic activity and factors V, VII and VIII. *Proc R Soc Med,* **68,** 267-268.

Chan, V.S., Chan, K.Y., Chen, Y., Poon, L.L., Cheung, A.N., Zheng, B., Chan, K.H., Mak, W., Ngan, H.Y., Xu, X., Screaton, G., Tam, P.K., Austyn, J.M., Chan, L.C., Yip, S.P., Peiris, M., Khoo, U.S. & Lin, C.L. (2006) Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. *Nat Genet,* **38,** 38-46.

Conlan, M.G., Folsom, A.R., Finch, A., Davis, C.E., Sorlie, P., Marcucci, G. & Wu, K.K. (1993) Associations of factor VIII and von Willebrand factor with age, race, sex, and risk factors for atherosclerosis. The Atherosclerosis Risk in Communities (ARIC) Study. *Thromb Haemost,* **70,** 380-385.

Corrales, I., Catarino, S., Ayats, J., Arteta, D., Altisent, C., Parra, R. & Vidal, F. (2012) High-throughput molecular diagnosis of von Willebrand disease by next generation sequencing methods. *Haematologica,* **97,** 1003-1007.

Costache, M., Cailleau, A., Fernandez-Mateos, P., Oriol, R. & Mollicone, R. (1997) Advances in molecular genetics of alpha-2- and alpha-3/4-fucosyltransferases. *Transfus Clin Biol,* **4,** 367-382.

Crawley, J.T.B., de Groot, R., Xiang, Y., Luken, B.M. & Lane, D.A. (2011) Unraveling the scissile bond: how ADAMTS13 recognizes and cleaves von Willebrand factor. *Blood,* **118,** 3212-3221.

Cumming, A., Grundy, P., Keeney, S., Lester, W., Enayat, S., Guilliatt, A., Bowen, D., Pasi, J., Keeling, D., Hill, F., Bolton-Maggs, P.H.B., Hay, C. & Collins, P. (2006) An investigation of the von Willebrand factor genotype in UK patients diagnosed to have type 1 von Willebrand disease. *Thromb Haemost,* **96,** 630-641.

Daidone, V., Cattini, M.G., Pontara, E., Sartorello, F., Gallinaro, L., Marotti, A., Scaroni, C., Pagnan, A. & Casonato, A. (2009) Microsatellite (GT)(n) repeats and SNPs in the von Willebrand factor gene promoter do not influence circulating von Willebrand factor levels under normal conditions. *Thromb Haemost,* **101,** 298-304.

Dall'Olio, F., Malagolini, N., Chiricolo, M., Trinchera, M. & Harduin-Lepers, A. (2014) The expanding roles of the Sd(a)/Cad carbohydrate antigen and its cognate glycosyltransferase B4GALNT2. *Biochim Biophys Acta,* **1840,** 443-453.

De Visser, M.C.H., Sandkuijl, L.A., Lensen, R.P.M., Vos, H.L., Rosendaal, F.R. & Bertina, R.M. (2003) Linkage analysis of factor VIII and von Willebrand factor loci as quantitative trait loci. *J Thromb Haemost,* **1,** 1771-1776.

Dean, L. (2005) Blood Groups and Red Cell Antigens In: *The ABO blood group*. Bethesda (MD) : National Center for Biotechnology Information, (US).

Denis, C.V., Christophe, O.D., Oortwijn, B.D. & Lenting, P.J. (2008) Clearance of von Willebrand factor. *Thromb Haemost,* **99,** 271-278.

Desch, K.C. (2012) Genetic modifiers of VWF antigen levels in healthy young adults. *University of Michigan,*

Ann Arbor, MI USA [abstract].

Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M. & Beroud, C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res,* **37,** e67.

Di Paola, J., Jugessur, A., Goldman, T., Reiland, J., Tallman, D., Sayago, C. & Murray, J.C. (2005) Platelet glycoprotein I(b)alpha and integrin alpha2 beta1 polymorphisms: gene frequencies and linkage disequilibrium in a population diversity panel. *J Thromb Haemost,* **3,** 1511-1521.

Dong, Z., Thoma, R.S., Crimmins, D.L., McCourt, D.W., Tuley, E.A. & Sadler, J.E. (1994) Disulfide bonds required to assemble functional von Willebrand factor multimers. *J Biol Chem,* **269,** 6753-6758.

Duan, J., Shi, J., Ge, X., Dolken, L., Moy, W., He, D., Shi, S., Sanders, A.R., Ross, J. & Gejman, P.V. (2013) Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. *Sci Rep,* **3,** 1318.

Eikenboom, J., Federici, A.B., Dirven, R.J., Castaman, G., Rodeghiero, F., Budde, U., Schneppenheim, R., Batlle, J., Canciani, M.T., Goudemand, J., Peake, I. & Goodeve, A. (2013) VWF propeptide and ratios between VWF, VWF propeptide, and FVIII in the characterization of type 1 von Willebrand disease. *Blood,* **121,** 2336-2339.

Eikenboom, J., Van Marion, V., Putter, H., Goodeve, A., Rodeghiero, F., Castaman, G., Federici, A.B., Batlle, J., Meyer, D., Mazurier, C., Goudemand, J., Schneppenheim, R., Budde, U., Ingerslev, J., Vorlova, Z., Habart, D., Holmberg, L., Lethagen, S., Pasi, J., Hill, F. & Peake, I. (2006) Linkage analysis in families diagnosed with type 1 von Willebrand disease in the European study, molecular and clinical markers for the diagnosis and management of type 1 VWD. *J Thromb Haemost,* **4,** 774-782.

Eikenboom, J.C., Matsushita, T., Reitsma, P.H., Tuley, E.A., Castaman, G., Briet, E. & Sadler, J.E. (1996) Dominant type 1 von Willebrand disease caused by mutated cysteine residues in the D3 domain of von Willebrand factor. *Blood,* **88,** 2433-2441.

Eikenboom, J.C., Vink, T., Briet, E., Sixma, J.J. & Reitsma, P.H. (1994) Multiple substitutions in the von Willebrand factor gene that mimic the pseudogene sequence. *Proc Natl Acad Sci U S A,* **91,** 2221-2224.

Ellies, L.G., Ditto, D., Levy, G.G., Wahrenbrock, M., Ginsburg, D., Varki, A., Le, D.T. & Marth, J.D. (2002) Sialyltransferase ST3Gal-IV operates as a dominant modifier of hemostasis by concealing asialoglycoprotein receptor ligands. *Proc Natl Acad Sci U S A,* **99,** 10042-10047.

Elmgren, A., Borjeson, C., Mollicone, R., Oriol, R., Fletcher, A. & Larson, G. (2000) Identification of two functionally deficient plasma alpha 3-fucosyltransferase (FUT6) alleles. *Hum Mutat,* **16,** 473-481.

Elmgren, A., Mollicone, R., Costache, M., Borjeson, C., Oriol, R., Harrington, J. & Larson, G. (1997) Significance of individual point mutations, T202C and C314T, in the human Lewis (FUT3) gene for expression of Lewis antigens by the human alpha(1,3/1,4)-fucosyltransferase, Fuc-TIII. *J Biol Chem,* **272,** 21994-21998.

Enayat, M.S., Guilliatt, A.M., Surdhar, G.K., Jenkins, P.V., Pasi, K.J., Toh, C.H., Williams, M.D. & Hill, F.G. (2001) Aberrant dimerization of von Willebrand factor as the result of mutations in the carboxy-terminal region: identification of 3 mutations in members of 3 different families with type 2A (phenotype IID) von Willebrand disease. *Blood,* **98,** 674-680.

Fabian, M.R., Sonenberg, N. & Filipowicz, W. (2010) Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem,* **79,** 351-379.

Favaloro, E.J., Soltani, S., McDonald, J., Grezchnik, E., Easton, L. & Favaloro, J.W.C. (2005) Reassessment of ABO blood group, sex, and age on laboratory parameters used to diagnose von Willebrand disorder: potential influence on the diagnosis vs the potential association with risk of thrombosis. *Am J Clin Pathol,* **124,** 910-917.

Feinberg, H., Guo, Y., Mitchell, D.A., Drickamer, K. & Weis, W.I. (2005) Extended neck regions stabilize tetramers of the receptors DC-SIGN and DC-SIGNR. *J Biol Chem,* **280,** 1327-1335.

Ferraro, F., Kriston-Vizi, J., Metcalf, D.J., Martin-Martin, B., Freeman, J., Burden, J.J., Westmoreland, D., Dyer, C.E., Knight, A.E., Ketteler, R. & Cutler, D.F. (2014) A two-tier Golgi-based control of organelle size underpins the functional plasticity of endothelial cells. *Dev Cell,* **29,** 292-304.

Fujimura, Y., Titani, K., Holland, L.Z., Russell, S.R., Roberts, J.R., Elder, J.H., Ruggeri, Z.M. & Zimmerman, T.S. (1986) von Willebrand factor. A reduced and alkylated 52/48-kDa fragment beginning at amino acid residue 449 contains the domain interacting with platelet glycoprotein Ib. *J Biol Chem,* **261,** 381-385.

Furlan, M., Robles, R. & Lammle, B. (1996) Partial purification and characterization of a protease from human plasma cleaving von Willebrand factor to fragments produced by in vivo proteolysis. *Blood,* **87,** 4223-4234.

Gadisseur, A., Berneman, Z., Schroyens, W. & Michiels, J.J. (2009) Laboratory diagnosis of von Willebrand disease type 1/2E (2A subtype IIE), type 1 Vicenza and mild type 1 caused by mutations in the D3, D4, B1-B3 and C1-C2 domains of the von Willebrand factor gene. Role of von Willebrand factor multimers and the von Willebrand factor propeptide/antigen ratio. *Acta Haematol,* **121,** 128-138.

Gaildrat, P., Krieger, S., Di Giacomo, D., Abdat, J., Revillion, F., Caputo, S., Vaur, D., Jamard, E., Bohers, E., Ledemeney, D., Peyrat, J.P., Houdayer, C., Rouleau, E., Lidereau, R., Frebourg, T., Hardouin, A., Tosi, M. & Martins, A. (2012) Multiple sequence variants of BRCA2 exon 7 alter splicing regulation. *J Med Genet,* **49,** 609-617.

Galbusera, M., Zoja, C., Donadelli, R., Paris, S., Morigi, M., Benigni, A., Figliuzzi, M., Remuzzi, G. & Remuzzi, A. (1997) Fluid shear stress modulates von Willebrand factor release from human vascular endothelium. *Blood,* **90,** 1558-1564.

Gallinaro, L., Cattini, M.G., Sztukowska, M., Padrini, R., Sartorello, F., Pontara, E., Bertomoro, A., Daidone, V., Pagnan, A. & Casonato, A. (2008) A shorter von Willebrand factor survival in O blood group

subjects explains how ABO determinants influence plasma von Willebrand factor. *Blood,* **111,** 3540-3545.

Gallinaro, L., Sartorello, F., Pontara, E., Cattini, M.G., Bertomoro, A., Bartoloni, L., Pagnan, A. & Casonato, A. (2006) Combined partial exon skipping and cryptic splice site activation as a new molecular mechanism for recessive type 1 von Willebrand disease. *Thromb Haemost,* **96,** 711-716.

Gill, J.C., Endres-Brooks, J., Bauer, P.J., Marks, W.J., Jr. & Montgomery, R.R. (1987) The effect of ABO blood group on the diagnosis of von Willebrand disease. *Blood,* **69,** 1691-1695.

Ginsburg, D., Handin, R.I., Bonthron, D.T., Donlon, T.A., Bruns, G.A., Latt, S.A. & Orkin, S.H. (1985) Human von Willebrand factor (vWF): isolation of complementary DNA (cDNA) clones and chromosomal localization. *Science,* **228,** 1401-1406.

Goodeve, A., Eikenboom, J., Castaman, G., Rodeghiero, F., Federici, A.B., Batlle, J., Meyer, D., Mazurier, C., Goudemand, J., Schneppenheim, R., Budde, U., Ingerslev, J., Habart, D., Vorlova, Z., Holmberg, L., Lethagen, S., Pasi, J., Hill, F., Hashemi Soteh, M., Baronciani, L., Hallden, C., Guilliatt, A., Lester, W. & Peake, I. (2007) Phenotype and genotype of a cohort of families historically diagnosed with type 1 von Willebrand disease in the European study, Molecular and Clinical Markers for the Diagnosis and Management of Type 1 von Willebrand Disease (MCMDM-1VWD). *Blood,* **109,** 112-121.

Goodeve, A.C. (2010) The genetic basis of von Willebrand disease. *Blood Rev,* **24,** 123-134.

Green, D., Jarrett, O., Ruth, K.J., Folsom, A.R. & Liu, K. (1995) Relationship among Lewis phenotype, clotting factors, and other cardiovascular risk factors in young adults. *J Lab Clin Med,* **125,** 334-339.

Grewal, P.K., Uchiyama, S., Ditto, D., Varki, N., Le, D.T., Nizet, V. & Marth, J.D. (2008) The Ashwell receptor mitigates the lethal coagulopathy of sepsis. *Nat Med,* **14,** 648-655.

Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res,* **36,** D154-158.

Guo, Y., Feinberg, H., Conroy, E., Mitchell, D.A., Alvarez, R., Blixt, O., Taylor, M.E., Weis, W.I. & Drickamer, K. (2004) Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat Struct Mol Biol,* **11,** 591-598.

Haberichter, S.L., Merricks, E.P., Fahs, S.A., Christopherson, P.A., Nichols, T.C. & Montgomery, R.R. (2005) Re-establishment of VWF-dependent Weibel-Palade bodies in VWD endothelial cells. *Blood,* **105,** 145-152.

Hampshire, D.J., Burghel, G.J., Goudemand, J., Bouvet, L.C., Eikenboom, J.C., Schneppenheim, R., Budde, U., Peake, I.R. & Goodeve, A.C. (2010) Polymorphic variation within the VWF gene contributes to the failure to detect mutations in patients historically diagnosed with type 1 von Willebrand disease from the MCMDM-1VWD cohort. *Haematologica,* **95,** 2163-2165.

Harrison, R.L. & McKee, P.A. (1984) Estrogen stimulates von Willebrand factor production by cultured endothelial cells. *Blood,* **63,** 657-664.

Harvey, P.J., Keightley, A.M., Lam, Y.M., Cameron, C. & Lillicrap, D. (2000) A single nucleotide polymorphism at nucleotide -1793 in the von Willebrand factor (VWF) regulatory region is associated with plasma VWF:Ag levels. *Br J Haematol,* **109,** 349-353.

Hickson, N., Hampshire, D., Castaman, G., Eikenboom, J., Rodeghiero, F., Peake, I. & Goodeve, A. (2011) Effect of the VWF promoter (GT)n repeat and single-nucleotide polymorphism c.-2527G>A on circulating von Willebrand factor levels under normal conditions. *J Thromb Haemost,* **9,** 603-605.

Hickson, N., Hampshire, D., Winship, P., Goudemand, J., Schneppenheim, R., Budde, U., Castaman, G., Rodeghiero, F., Federici, A.B., James, P., Peake, I., Eikenboom, J. & Goodeve, A. (2010) von Willebrand factor variant p.Arg924Gln marks an allele associated with reduced von Willebrand factor and factor VIII levels. *J Thromb Haemost,* **8,** 1986-1993.

Hickson, N., Hampshire D.J., Winship P.R., James P.D., Peake I.R., Goodeve A.C., EU-VWD & groups, a.Z.-V.s. (2009) FUT3 gene determining Lewis blood group antigen expression influences VWF:Ag levels in plasma. *J THROMB HAEMOST, Vol. 7(Suppl. 2) (pp 1147-1147)*.

Hilbert, L., Jorieux, S., Proulle, V., Favier, R., Goudemand, J., Parquet, A., Meyer, D., Fressinaud, E. & Mazurier, C. (2003) Two novel mutations, Q1053H and C1060R, located in the D3 domain of von Willebrand factor, are responsible for decreased FVIII-binding capacity. *Br J Haematol,* **120,** 627-632.

Hough, C., Cameron, C.L., Notley, C.R., Brown, C., O'Brien, L., Keightley, A.M., Berber, E. & Lillicrap, D. (2008) Influence of a GT repeat element on shear stress responsiveness of the VWF gene promoter. *J Thromb Haemost,* **6,** 1183-1190.

Hoylaerts, M.F., Yamamoto, H., Nuyts, K., Vreys, I., Deckmyn, H. & Vermylen, J. (1997) von Willebrand factor binds to native collagen VI primarily via its A1 domain. *Biochem J,* **324 (Pt 1),** 185-191.

Iverson, C., Christiansen, S., Flanagin, A. & al, e. (2007) AMA Manual of Style: A Guide for Authors and Editors. . *New York, NY: Oxford University Press. © American Medical Association,* **10th ed.**

Jaffe, E.A., Hoyer, L.W. & Nachman, R.L. (1973) Synthesis of antihemophilic factor antigen by cultured human endothelial cells. *J Clin Invest,* **52,** 2757-2764.

James, P.D., Notley, C., Hegadorn, C., Leggo, J., Tuttle, A., Tinlin, S., Brown, C., Andrews, C., Labelle, A., Chirinian, Y., O'Brien, L., Othman, M., Rivard, G., Rapson, D., Hough, C. & Lillicrap, D. (2007) The mutational spectrum of type 1 von Willebrand disease: Results from a Canadian cohort study. *Blood,* **109,** 145-154.

Jansson, J.H., Nilsson, T.K. & Johnson, O. (1991) von Willebrand factor in plasma: a novel risk factor for recurrent myocardial infarction and death. *Br Heart J,* **66,** 351-355.

Jenkins, P.V., Rawley, O., Smith, O.P. & O'Donnell, J.S. (2012) Elevated factor VIII levels and risk of venous thrombosis. *Br J Haematol,* **157,** 653-663.

Johnsen, J.M., Auer, P.L., Morrison, A.C., Jiao, S., Wei, P., Haessler, J., Fox, K., McGee, S.R., Smith, J.D., Carlson, C.S., Smith, N., Boerwinkle, E., Kooperberg, C., Nickerson, D.A., Rich, S.S., Green, D., Peters, U., Cushman, M. & Reiner, A.P. (2013) Common and rare von Willebrand factor (VWF) coding variants, VWF levels, and factor VIII levels in African Americans: the NHLBI Exome Sequencing Project. *Blood,* **122,** 590-597.

Kadir, R.A., Economides, D.L., Sabin, C.A., Owens, D. & Lee, C.A. (1999) Variations in coagulation factors in women: effects of age, ethnicity, menstrual cycle and combined oral contraceptive. *Thromb Haemost,* **82,** 1456-1461.

Katsumi, A., Tuley, E.A., Bodo, I. & Sadler, J.E. (2000) Localization of disulfide bonds in the cystine knot domain of human von Willebrand factor. *J Biol Chem,* **275,** 25585-25594.

Kaufmann, J.E., Oksche, A., Wollheim, C.B., Gunther, G., Rosenthal, W. & Vischer, U.M. (2000) Vasopressin-induced von Willebrand factor secretion from endothelial cells involves V2 receptors and cAMP. *J Clin Invest,* **106,** 107-116.

Keeney, S. & Cumming, A.M. (2001) The molecular biology of von Willebrand disease. *Clin Lab Haematol,* **23,** 209-230.

Keightley, A.M., Lam, Y.M., Brady, J.N., Cameron, C.L. & Lillicrap, D. (1999) Variation at the von Willebrand factor (vWF) gene locus is associated with plasma vWF:Ag levels: identification of three novel single nucleotide polymorphisms in the vWF gene promoter. *Blood,* **93,** 4277-4283.

Kenny, D., Morateck, P.A. & Montgomery, R.R. (2002) The cysteine knot of platelet glycoprotein Ib beta (GPIb beta) is critical for the interaction of GPIb beta with GPIX. *Blood,* **99,** 4428-4433.

Khoo, U.S., Chan, K.Y., Chan, V.S. & Lin, C.L. (2008) DC-SIGN and L-SIGN: the SIGNs for infection. *J Mol Med (Berl),* **86,** 861-874.

Klarmann, D., Eggert, C., Geisen, C., Becker, S., Seifried, E., Klingebiel, T. & Kreuz, W. (2010) Association of ABO(H) and I blood group system development with von Willebrand factor and Factor VIII plasma levels in children and adolescents. *Transfusion,* **50,** 1571-1580.

Konkle, B.A. (2014) Von Willebrand factor and aging. *Semin Thromb Hemost,* **40,** 640-644.

Kozomara, A. & Griffiths-Jones, S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res,* **39,** D152-157.

Kroner, P.A., Friedman, K.D., Fahs, S.A., Scott, J.P. & Montgomery, R.R. (1991) Abnormal binding of factor VIII is linked with the substitution of glutamine for arginine 91 in von Willebrand factor in a variant form of von Willebrand disease. *J Biol Chem,* **266,** 19146-19149.

Ky, B., French, B., Ruparel, K., Sweitzer, N.K., Fang, J.C., Levy, W.C., Sawyer, D.B. & Cappola, T.P. (2011) The vascular marker soluble fms-like tyrosine kinase 1 is associated with disease severity and adverse outcomes in chronic heart failure. *J Am Coll Cardiol,* **58,** 386-394.

Lacquemant, C., Gaucher, C., Delorme, C., Chatellier, G., Gallois, Y., Rodier, M., Passa, P., Balkau, B., Mazurier, C., Marre, M. & Froguel, P. (2000) Association between high von willebrand factor levels and the Thr789Ala vWF gene polymorphism but not with nephropathy in type I diabetes. The GENEDIAB Study Group and the DESIR Study Group. *Kidney Int,* **57,** 1437-1443.

Laffan, M., Brown, S.A., Collins, P.W., Cumming, A.M., Hill, F.G., Keeling, D., Peake, I.R. & Pasi, K.J. (2004) The diagnosis of von Willebrand disease: a guideline from the UK Haemophilia Centre Doctors' Organization. *Haemophilia,* **10,** 199-217.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics,* **23,** 2947-2948.

Leclerc, G.J., Leclerc, G.M. & Barredo, J.C. (2002) Real-time RT-PCR analysis of mRNA decay: half-life of Beta-actin mRNA in human leukemia CCRF-CEM and Nalm-6 cell lines. *Cancer Cell Int,* **2,** 1.

Lemmerhirt, H.L., Broman, K.W., Shavit, J.A. & Ginsburg, D. (2007) Genetic regulation of plasma von Willebrand factor levels: quantitative trait loci analysis in a mouse model. *J Thromb Haemost,* **5,** 329-335.

Lemmerhirt, H.L., Shavit, J.A., Levy, G.G., Cole, S.M., Long, J.C. & Ginsburg, D. (2006) Enhanced VWF biosynthesis and elevated plasma VWF due to a natural variant in the murine Vwf gene. *Blood,* **108,** 3061-3067.

Lenting, P.J., Casari, C., Christophe, O.D. & Denis, C.V. (2012) Von Willebrand factor: the old, the new and the unknown. *J Thromb Haemost*.

Lenting, P.J., Christophe, O.D. & Denis, C.V. (2015) von Willebrand factor biosynthesis, secretion, and clearance: connecting the far ends. *Blood,* **125,** 2019-2028.

Lenting, P.J., Pegon, J.N., Christophe, O.D. & Denis, C.V. (2010) Factor VIII and von Willebrand factor--too sweet for their own good. *Haemophilia,* **16 Suppl 5,** 194-199.

Lester, W., Guilliatt, A., Grundy, P., Enayat, S., Millar, C., Hill, F., Cumming, T. & Collins, P. (2008) Is VWF R924Q a benign polymorphism, a marker of a null allele or a factor VIII-binding defect? The debate continues with results from the UKHCDO VWD study. *Thromb Haemost,* **100,** 716-718.

Lethagen, S., Hillarp, A., Ekholm, C., Mattson, E., Hallden, C. & Friberg, B. (2008) Distribution of von Willebrand factor levels in young women with and without bleeding symptoms: influence of ABO blood group and promoter haplotypes. *Thromb Haemost,* **99,** 1013-1018.

Li, H., Wang, J.X., Wu, D.D., Wang, H.W., Tang, N.L. & Zhang, Y.P. (2012) The Origin and Evolution of Variable Number Tandem Repeat of CLEC4M Gene in the Global Human Population. *PLoS One,* **7,** e30268.

Lillicrap, D. (2007) Von Willebrand disease - phenotype versus genotype: deficiency versus disease. *Thromb Res,* **120 Suppl 1,** S11-16.

Livak, K.J. & Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* **25,** 402-408.

Lopez, J.A., Ludwig, E.H. & McCarthy, B.J. (1992) Polymorphism of human glycoprotein Ib alpha results from a variable number of tandem repeats of a 13-amino acid sequence in the mucin-like macroglycopeptide region. Structure/function implications. *J Biol Chem,* **267,** 10055-10061.

Lynch, D.C., Zimmerman, T.S., Ling, E.H. & Browning, P.J. (1986) An explanation for minor multimer species in endothelial cell-synthesized von Willebrand factor. *J Clin Invest,* **77,** 2048-2051.

Mancuso, D.J., Tuley, E.A., Westfield, L.A., Lester-Mancuso, T.L., Le Beau, M.M., Sorace, J.M. & Sadler, J.E. (1991) Human von Willebrand factor

gene and pseudogene: structural analysis and differentiation by polymerase chain reaction. *Biochemistry,* **30,** 253-269.

Mancuso, D.J., Tuley, E.A., Westfield, L.A., Worrall, N.K., Shelton-Inloes, B.B., Sorace, J.M., Alevy, Y.G. & Sadler, J.E. (1989) Structure of the gene for human von Willebrand factor. *J Biol Chem,* **264,** 19514-19527.

Mannucci, P.M., Capoferri, C. & Canciani, M.T. (2004) Plasma levels of von Willebrand factor regulate ADAMTS-13, its major cleaving protease. *Br J Haematol,* **126,** 213-218.

Mannucci, P.M., Lattuada, A., Castaman, G., Lombardi, R., Colibretti, M.L., Ciavarella, N. & Rodeghiero, F. (1989) Heterogeneous phenotypes of platelet and plasma von Willebrand factor in obligatory heterozygotes for severe von Willebrand disease. *Blood,* **74,** 2433-2436.

Mathe, E., Olivier, M., Kato, S., Ishioka, C., Hainaut, P. & Tavtigian, S.V. (2006) Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res,* **34,** 1317-1325.

Matsui, T., Fujimura, Y., Nishida, S. & Titani, K. (1993) Human plasma alpha 2-macroglobulin and von Willebrand factor possess covalently linked ABO(H) blood group antigens in subjects with corresponding ABO phenotype. *Blood,* **82,** 663-668.

Matsui, T., Titani, K. & Mizuochi, T. (1992) Structures of the asparagine-linked oligosaccharide chains of human von Willebrand factor. Occurrence of blood group A, B, and H(O) structures. *J Biol Chem,* **267,** 8723-8731.

Mazzeffi, M.A. & Stone, M.E. (2011) Perioperative management of von Willebrand disease: a review for the anesthesiologist. *J Clin Anesth,* **23,** 418-426.

Mazzucato, M., Spessotto, P., Masotti, A., De Appollonia, L., Cozzi, M.R., Yoshioka, A., Perris, R., Colombatti, A. & De Marco, L. (1999) Identification of domains responsible for von Willebrand factor type VI collagen interaction mediating platelet adhesion under high flow. *J Biol Chem,* **274,** 3033-3041.

McGrath, R.T., McRae, E., Smith, O.P. & O'Donnell, J.S. (2010) Platelet von Willebrand factor--structure, function and biological importance. *Br J Haematol,* **148,** 834-843.

McKinnon, T.A.J., Chion, A.C.K., Millington, A.J., Lane, D.A. & Laffan, M.A. (2008) N-linked glycosylation of VWF modulates its interaction with ADAMTS13. *Blood,* **111,** 3042-3049.

McKinnon, T.A.J., Goode, E.C., Birdsey, G.M., Nowak, A.A., Chan, A.C.K., Lane, D.A. & Laffan, M.A. (2010) Specific N-linked glycosylation sites

modulate synthesis and secretion of von Willebrand factor. *Blood,* **116,** 640-648.

Meade, T.W., Brozovic, M., Chakrabarti, R., Haines, A.P., North, W.R. & Stirling, Y. (1978) Ethnic group comparisons of variables associated with ischaemic heart disease. *Br Heart J,* **40,** 789-795.

Meade, T.W., North, W.R., Chakrabarti, R., Haines, A.P. & Stirling, Y. (1977) Population-based distributions of haemostatic variables. *Br Med Bull,* **33,** 283-288.

Mello, T.B., Siqueira, L.H., Montavao, S.A., Ozello, M.C. & Annichino-Bizzacchi, J.M. (2008) Low density lipoprotein receptor-related protein polymorphisms are not risk factors for venous thromboembolism. *Thromb Res,* **121,** 625-629.

Mercier, B., Gaucher, C. & Mazurier, C. (1991) Characterisation of 98 alleles in 105 unrelated individuals in the F8VWF gene. *Nucleic Acids Res,* **19,** 4800.

Millar, C.M. & Brown, S.A. (2006) Oligosaccharide structures of von Willebrand factor and their potential role in von Willebrand disease. *Blood Rev,* **20,** 83-92.

Mitchell, D.A., Fadden, A.J. & Drickamer, K. (2001) A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J Biol Chem,* **276,** 28939-28945.

Mohlke, K.L., Nichols, W.C., Westrick, R.J., Novak, E.K., Cooney, K.A., Swank, R.T. & Ginsburg, D. (1996) A novel modifier gene for plasma von Willebrand factor level maps to distal mouse chromosome 11. *Proc Natl Acad Sci U S A,* **93,** 15352-15357.

Mohlke, K.L., Purkayastha, A.A., Westrick, R.J., Smith, P.L., Petryniak, B., Lowe, J.B. & Ginsburg, D. (1999) Mvwf, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosyltransferase. *Cell,* **96,** 111-120.

Mollicone, R., Cailleau, A. & Oriol, R. (1995) Molecular genetics of H, Se, Lewis and other fucosyltransferase genes. *Transfus Clin Biol,* **2,** 235-242.

Mollicone, R., Reguigne, I., Fletcher, A., Aziz, A., Rustam, M., Weston, B.W., Kelly, R.J., Lowe, J.B. & Oriol, R. (1994) Molecular basis for plasma alpha(1,3)-fucosyltransferase gene deficiency (FUT6). *J Biol Chem,* **269,** 12662-12671.

Montgomery, R.R., Hathaway, W.E., Johnson, J., Jacobson, L. & Muntean, W. (1982) A variant of von Willebrand's disease with abnormal expression of factor VIII procoagulant activity. *Blood,* **60,** 201-207.

Morange, P.E., Saut, N., Antoni, G., Emmerich, J. & Tregouet, D.A. (2011) Impact on venous thrombosis risk of newly discovered gene variants associated with FVIII and VWF plasma levels. *J Thromb Haemost,* **9,** 229-231.

Morange, P.E., Suchon, P. & Tregouet, D.A. (2015) Genetics of Venous Thrombosis: update in 2015. *Thromb Haemost,* **114,** 910-919.

Morange, P.E., Tregouet, D.A., Frere, C., Saut, N., Pellegrina, L., Alessi, M.C., Visvikis, S., Tiret, L. & Juhan-Vague, I. (2005) Biological and genetic factors influencing plasma factor VIII levels in a healthy family population: results from the Stanislas cohort. *Br J Haematol,* **128,** 91-99.

Morga-Ramirez, M., Collados-Larumbe, M.T., Johnson, K.E., Rivas-Arreola, M.J., Carrillo-Cocom, L.M. & Alvarez, M.M. (2010) Hydrodynamic conditions induce changes in secretion level and glycosylation patterns of Von Willebrand factor (vWF) in endothelial cells. *J Biosci Bioeng,* **109,** 400-406.

Mullis, K.B. (1990) The unusual origin of the polymerase chain reaction. *Sci Am,* **262,** 56-61, 64-55.

Nichols, W.C., Cooney, K.A., Mohlke, K.L., Ballew, J.D., Yang, A., Bruck, M.E., Reddington, M., Novak, E.K., Swank, R.T. & Ginsburg, D. (1994) von Willebrand disease in the RIIIS/J mouse is caused by a defect outside of the von Willebrand factor gene. *Blood,* **83,** 3225-3231.

Nichols, W.L., Hultin, M.B., James, A.H., Manco-Johnson, M.J., Montgomery, R.R., Ortel, T.L., Rick, M.E., Sadler, J.E., Weinstein, M. & Yawn, B.P. (2008) von Willebrand disease (VWD): evidence-based diagnosis and management guidelines, the National Heart, Lung, and Blood Institute (NHLBI) Expert Panel report (USA). *Haemophilia,* **14,** 171-232.

Nightingale, T. & Cutler, D. (2013) The secretion of von Willebrand factor from endothelial cells; an increasingly complicated story. *J Thromb Haemost,* **11 Suppl 1,** 192-201.

Nishino, M., Girma, J.P., Rothschild, C., Fressinaud, E. & Meyer, D. (1989) New variant of von Willebrand disease with defective binding to factor VIII. *Blood,* **74,** 1591-1599.

Nossent, A.Y., Robben, J.H., Deen, P.M.T., Vos, H.L., Rosendaal, F.R., Doggen, C.J.M., Hansen, J.L., Sheikh, S.P., Bertina, R.M. & Eikenboom, J.C.J. (2010) Functional variation in the arginine vasopressin 2 receptor as a modifier of human plasma von Willebrand factor levels. *J Thromb Haemost,* **8,** 1547-1554.

Nuzzo, F., Bulato, C., Nielsen, B.I., Lee, K., Wielders, S.J., Simioni, P., Key, N.S. & Castoldi, E. (2015) Characterization of an apparently

synonymous F5 mutation causing aberrant splicing and factor V deficiency. *Haemophilia,* **21,** 241-248.

O'Donnell, J., Boulton, F.E., Manning, R.A. & Laffan, M.A. (2002) Genotype at the secretor blood group locus is a determinant of plasma von Willebrand factor level. *Br J Haematol,* **116,** 350-356.

O'Donnell, J.S., McKinnon, T.A.J., Crawley, J.T.B., Lane, D.A. & Laffan, M.A. (2005) Bombay phenotype is associated with reduced plasma-VWF levels and an increased susceptibility to ADAMTS13 proteolysis. *Blood,* **106,** 1988-1991.

Orstavik, K.H., Kornstad, L., Reisner, H. & Berg, K. (1989) Possible effect of secretor locus on plasma concentration of factor VIII and von Willebrand factor. *Blood,* **73,** 990-993.

Pang, H., Koda, Y., Soejima, M., Schlaphoff, T., du Toit, E.D. & Kimura, H. (1999) Allelic diversity of the human plasma alpha(1,3)fucosyltransferase gene (FUT6). *Ann Hum Genet,* **63,** 277-284.

Pegon, J.N., Kurdi, M., Casari, C., Odouard, S., Denis, C.V., Christophe, O.D. & Lenting, P.J. (2012) Factor VIII and von Willebrand factor are ligands for the carbohydrate-receptor Siglec-5. *Haematologica,* **97,** 1855-1863.

Rastegarlari, G., Pegon, J.N., Casari, C., Odouard, S., Navarrete, A.M., Saint-Lu, N., van Vlijmen, B.J., Legendre, P., Christophe, O.D., Denis, C.V. & Lenting, P.J. (2012) Macrophage LRP1 contributes to the clearance of von Willebrand factor. *Blood*.

Reese, M.G., Eeckman, F.H., Kulp, D. & Haussler, D. (1997) Improved splice site detection in Genie. *J Comput Biol,* **4,** 311-323.

Reininger, A.J. (2008) Function of von Willebrand factor in haemostasis and thrombosis. *Haemophilia,* **14 Suppl 5,** 11-26.

Rodeghiero, F., Castaman, G. & Dini, E. (1987) Epidemiological investigation of the prevalence of von Willebrand's disease. *Blood,* **69,** 454-459.

Rondaij, M.G., Bierings, R., Kragt, A., van Mourik, J.A. & Voorberg, J. (2006) Dynamics and plasticity of Weibel-Palade bodies in endothelial cells. *Arterioscler Thromb Vasc Biol,* **26,** 1002-1007.

Rozen, S. & Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol,* **132,** 365-386.

Ruggeri, Z.M. (2001) Structure of von Willebrand factor and its function in platelet adhesion and thrombus formation. *Best Pract Res Clin Haematol,* **14,** 257-279.

Ruggeri, Z.M. & Zimmerman, T.S. (1980) Variant von Willebrand's disease: characterization of two subtypes by analysis of multimeric composition

of factor VIII/von Willebrand factor in plasma and platelets. *J Clin Invest,* **65,** 1318-1325.

Rydz, N., Boonyawat, B., Riches, J.J., Hopman, W., Paterson, A., Montgomery, R., Lillicrap, D., James, P.D. & Group5), a.t.Z.-V.S. (2011) Genetic Variability of the CLEC4M endothelial lectin receptor in type 1 von Willebrand disease.

Rydz, N., Swystun, L.L., Notley, C., Paterson, A.D., Riches, J.J., Sponagle, K., Boonyawat, B., Montgomery, R.R., James, P.D. & Lillicrap, D. (2013) The C-type lectin receptor CLEC4M binds, internalizes, and clears von Willebrand factor and contributes to the variation in plasma von Willebrand factor levels. *Blood,* **121,** 5228-5237.

Sabarinathan, R., Tafer, H., Seemann, S.E., Hofacker, I.L., Stadler, P.F. & Gorodkin, J. (2013) RNAsnp: efficient detection of local RNA secondary structure changes induced by SNPs. *Hum Mutat,* **34,** 546-556.

Sadler, J.E. (1998) Biochemistry and genetics of von Willebrand factor. *Annu Rev Biochem,* **67,** 395-424.

Sadler, J.E. (2004) Slippery criteria for von Willebrand disease type 1. *J Thromb Haemost,* **2,** 1720-1723.

Sadler, J.E. (2009) von Willebrand factor assembly and secretion. *J Thromb Haemost,* **7 Suppl 1,** 24-27.

Sadler, J.E., Budde, U., Eikenboom, J.C.J., Favaloro, E.J., Hill, F.G.H., Holmberg, L., Ingerslev, J., Lee, C.A., Lillicrap, D., Mannucci, P.M., Mazurier, C., Meyer, D., Nichols, W.L., Nishino, M., Peake, I.R., Rodeghiero, F., Schneppenheim, R., Ruggeri, Z.M., Srivastava, A., Montgomery, R.R. & Federici, A.B. (2006) Update on the pathophysiology and classification of von Willebrand disease: a report of the Subcommittee on von Willebrand Factor. *J Thromb Haemost,* **4,** 2103-2114.

Sadler, J.E., Shelton-Inloes, B.B., Sorace, J.M., Harlan, J.M., Titani, K. & Davie, E.W. (1985) Cloning and characterization of two cDNAs coding for human von Willebrand factor. *Proc Natl Acad Sci U S A,* **82,** 6394-6398.

Salomaa, V., Pankow, J., Heiss, G., Cakir, B., Eckfeldt, J.H., Ellison, R.C., Myers, R.H., Hiller, K.M., Brantley, K.R., Morris, T.L. & Weston, B.W. (2000) Genetic background of Lewis negative blood group phenotype and its association with atherosclerotic disease in the NHLBI family heart study. *J Intern Med,* **247,** 689-698.

Sanders, Y.V., van der Bom, J.G., Isaacs, A., Cnossen, M.H., de Maat, M.P., Laros-van Gorkom, B.A., Fijnvandraat, K., Meijer, K., van Duijn, C.M., Mauser-Bunschoten, E.P., Eikenboom, J. & Leebeek, F.W. (2015) CLEC4M and STXBP5 gene variations contribute to von Willebrand

factor level variation in von Willebrand disease. *J Thromb Haemost,* **13,** 956-966.

Savage, B., Saldivar, E. & Ruggeri, Z.M. (1996) Initiation of platelet adhesion by arrest onto fibrinogen or translocation on von Willebrand factor. *Cell,* **84,** 289-297.

Schneppenheim, R., Budde, U., Obser, T., Brassard, J., Mainusch, K., Ruggeri, Z.M., Schneppenheim, S., Schwaab, R. & Oldenburg, J. (2001) Expression and characterization of von Willebrand factor dimerization defects in different types of von Willebrand disease. *Blood,* **97,** 2059-2066.

Schwarz, J.M., Rodelsperger, C., Schuelke, M. & Seelow, D. (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods,* **7,** 575-576.

Shahbazi, S., Alavi, S. & Mahdian, R. (2012) Classification of exon 18 linked variants of VWF gene in von Willebrand disease. *Int J Mol Epidemiol Genet,* **3,** 77-83.

Shavit, J.A., Manichaikul, A., Lemmerhirt, H.L., Broman, K.W. & Ginsburg, D. (2009) Modifiers of von Willebrand factor identified by natural variation in inbred strains of mice. *Blood,* **114,** 5368-5374.

Shelton-Inloes, B.B., Titani, K. & Sadler, J.E. (1986) cDNA sequences for human von Willebrand factor reveal five types of repeated domains and five possible protein sequence polymorphisms. *Biochemistry,* **25,** 3164-3171.

Smith, N.L., Chen, M.-H., Dehghan, A., Strachan, D.P., Basu, S., Soranzo, N., Hayward, C., Rudan, I., Sabater-Lleal, M., Bis, J.C., de Maat, M.P.M., Rumley, A., Kong, X., Yang, Q., Williams, F.M.K., Vitart, V., Campbell, H., Malarstig, A., Wiggins, K.L., Van Duijn, C.M., McArdle, W.L., Pankow, J.S., Johnson, A.D., Silveira, A., McKnight, B., Uitterlinden, A.G., Aleksic, N., Meigs, J.B., Peters, A., Koenig, W., Cushman, M., Kathiresan, S., Rotter, J.I., Bovill, E.G., Hofman, A., Boerwinkle, E., Tofler, G.H., Peden, J.F., Psaty, B.M., Leebeek, F., Folsom, A.R., Larson, M.G., Spector, T.D., Wright, A.F., Wilson, J.F., Hamsten, A., Lumley, T., Witteman, J.C.M., Tang, W. & O'Donnell, C.J. (2010) Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation,* **121,** 1382-1392.

Soejima, M., Munkhtulga, L., Iwamoto, S. & Koda, Y. (2009) Genetic variation of FUT3 in Ghanaians, Caucasians, and Mongolians. *Transfusion,* **49,** 959-966.

Soilleux, E.J., Barten, R. & Trowsdale, J. (2000) DC-SIGN; a related gene, DC-SIGNR; and CD23 form a cluster on 19p13. *J Immunol,* **165,** 2937-2942.

Sole, X., Guino, E., Valls, J., Iniesta, R. & Moreno, V. (2006) SNPStats: a web tool for the analysis of association studies. *Bioinformatics,* **22,** 1928-1929.

Sorvillo, N., Kaijen, P.H., Matsumoto, M., Fujimura, Y., van der Zwaan, C., Verbij, F.C., Pos, W., Fijnheer, R., Voorberg, J. & Meijer, A.B. (2014) Identification of N-linked glycosylation and putative O-fucosylation, C-mannosylation sites in plasma derived ADAMTS13. *J Thromb Haemost,* **12,** 670-679.

Souto, J.C., Almasy, L., Soria, J.M., Buil, A., Stone, W., Lathrop, M., Blangero, J. & Fontcuberta, J. (2003) Genome-wide linkage analysis of von Willebrand factor plasma levels: results from the GAIT project. *Thromb Haemost,* **89,** 468-474.

Sporn, L.A., Chavin, S.I., Marder, V.J. & Wagner, D.D. (1985) Biosynthesis of von Willebrand protein by human megakaryocytes. *J Clin Invest,* **76,** 1102-1106.

Tang, W., Cushman, M., Green, D., Rich, S.S., Lange, L.A., Yang, Q., Tracy, R.P., Tofler, G.H., Basu, S., Wilson, J.G., Keating, B.J., Weng, L.C., Taylor, H.A., Jacobs, D.R., Jr., Delaney, J.A., Palmer, C.D., Young, T., Pankow, J.S., O'Donnell, C.J., Smith, N.L., Reiner, A.P. & Folsom, A.R. (2015) Gene-centric approach identifies new and known loci for FVIII activity and VWF antigen levels in European Americans and African Americans. *Am J Hematol,* **90,** 534-540.

Thompson, S.G., Kienast, J., Pyke, S.D., Haverkate, F. & van de Loo, J.C. (1995) Hemostatic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris. European Concerted Action on Thrombosis and Disabilities Angina Pectoris Study Group. *N Engl J Med,* **332,** 635-641.

Titani, K., Kumar, S., Takio, K., Ericsson, L.H., Wade, R.D., Ashida, K., Walsh, K.A., Chopek, M.W., Sadler, J.E. & Fujikawa, K. (1986) Amino acid sequence of human von Willebrand factor. *Biochemistry,* **25,** 3171-3184.

Tosetto, A., Rodeghiero, F., Castaman, G., Goodeve, A., Federici, A.B., Batlle, J., Meyer, D., Fressinaud, E., Mazurier, C., Goudemand, J., Eikenboom, J., Schneppenheim, R., Budde, U., Ingerslev, J., Vorlova, Z., Habart, D., Holmberg, L., Lethagen, S., Pasi, J., Hill, F. & Peake, I. (2006) A quantitative analysis of bleeding symptoms in type 1 von Willebrand disease: results from a multicenter European study (MCMDM-1 VWD). *J Thromb Haemost,* **4,** 766-773.

Tournier, I., Vezain, M., Martins, A., Charbonnier, F., Baert-Desurmont, S., Olschwang, S., Wang, Q., Buisine, M.P., Soret, J., Tazi, J., Frebourg, T. & Tosi, M. (2008) A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat,* **29,** 1412-1424.

Valencia-Sanchez, M.A., Liu, J., Hannon, G.J. & Parker, R. (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev,* **20,** 515-524.

Valentijn, K.M., Sadler, J.E., Valentijn, J.A., Voorberg, J. & Eikenboom, J. (2011) Functional architecture of Weibel-Palade bodies. *Blood,* **117,** 5033-5043.

Valentijn, K.M., Valentijn, J.A., Jansen, K.A. & Koster, A.J. (2008) A new look at Weibel-Palade body structure in endothelial cells using electron tomography. *J Struct Biol,* **161,** 447-458.

van Genderen, P.J., Prins, F.J., Lucas, I.S., van de Moesdijk, D., van Vliet, H.H., van Strik, R. & Michiels, J.J. (1997) Decreased half-life time of plasma von Willebrand factor collagen binding activity in essential thrombocythaemia: normalization after cytoreduction of the increased platelet count. *Br J Haematol,* **99,** 832-836.

van Schie, M.C., de Maat, M.P.M., Isaacs, A., van Duijn, C.M., Deckers, J.W., Dippel, D.W.J. & Leebeek, F.W.G. (2011a) Variation in the von Willebrand factor gene is associated with von Willebrand factor levels and with the risk for cardiovascular disease. *Blood,* **117,** 1393-1399.

van Schie, M.C., van Loon, J.E., de Maat, M.P. & Leebeek, F.W. (2011d) Genetic determinants of von Willebrand factor levels and activity in relation to the risk of cardiovascular disease: a review. *J Thromb Haemost,* **9,** 899-908.

van Schooten, C.J., Shahbazi, S., Groot, E., Oortwijn, B.D., van den Berg, H.M., Denis, C.V. & Lenting, P.J. (2008) Macrophages contribute to the cellular uptake of von Willebrand factor and factor VIII in vivo. *Blood,* **112,** 1704-1712.

van Schooten, C.J.M., Denis, C.V., Lisman, T., Eikenboom, J.C.J., Leebeek, F.W., Goudemand, J., Fressinaud, E., van den Berg, H.M., de Groot, P.G. & Lenting, P.J. (2007) Variations in glycosylation of von Willebrand factor with O-linked sialylated T antigen are associated with its plasma levels. *Blood,* **109,** 2430-2437.

Verweij, C.L., de Vries, C.J., Distel, B., van Zonneveld, A.J., van Kessel, A.G., van Mourik, J.A. & Pannekoek, H. (1985) Construction of cDNA coding for human von Willebrand factor using antibody probes for colony-screening and mapping of the chromosomal gene. *Nucleic Acids Res,* **13,** 4699-4717.

Wagner, D.D. (1990) Cell biology of von Willebrand factor. *Annu Rev Cell Biol,* **6,** 217-246.

Wagner, D.D., Mayadas, T. & Marder, V.J. (1986) Initial glycosylation and acidic pH in the Golgi apparatus are required for multimerization of von Willebrand factor. *J Cell Biol,* **102,** 1320-1324.

Wagner, D.D., Olmsted, J.B. & Marder, V.J. (1982) Immunolocalization of von Willebrand protein in Weibel-Palade bodies of human endothelial cells. *J Cell Biol,* **95,** 355-360.

Wang, M. & Marin, A. (2006) Characterization and prediction of alternative splice sites. *Gene,* **366,** 219-227.

Wohner, N., Legendre, P., Casari, C., Christophe, O.D., Lenting, P.J. & Denis, C.V. (2015) Shear stress-independent binding of von Willebrand factor-type 2B mutants p.R1306Q & p.V1316M to LRP1 explains their increased clearance. *J Thromb Haemost,* **13,** 815-820.

Yadegari, H., Driesen, J., Pavlova, A., Biswas, A., Ivaskevicius, V., Klamroth, R. & Oldenburg, J. (2013) Insights into pathological mechanisms of missense mutations in C-terminal domains of von Willebrand factor causing qualitative or quantitative von Willebrand disease. *Haematologica,* **98,** 1315-1323.

Yarovoi, H.V., Kufrin, D., Eslin, D.E., Thornton, M.A., Haberichter, S.L., Shi, Q., Zhu, H., Camire, R., Fakharzadeh, S.S., Kowalska, M.A., Wilcox, D.A., Sachais, B.S., Montgomery, R.R. & Poncz, M. (2003) Factor VIII ectopically expressed in platelets: efficacy in hemophilia A treatment. *Blood,* **102,** 4006-4013.

Yu, Q.D., Oldring, A.P., Powlesland, A.S., Tso, C.K., Yang, C., Drickamer, K. & Taylor, M.E. (2009) Autonomous tetramerization domains in the glycan-binding receptors DC-SIGN and DC-SIGNR. *J Mol Biol,* **387,** 1075-1080.

Zabaneh, D., Gaunt, T.R., Kumari, M., Drenos, F., Shah, S., Berry, D., Power, C., Hypponen, E., Shah, T., Palmen, J., Pallas, J., Talmud, P.J., Casas, J.P., Sofat, R., Lowe, G., Rumley, A., Morris, R.W., Whincup, P.H., Rodriguez, S., Ebrahim, S., Marmot, M.G., Smith, G.D., Lawlor, D.A., Kivimaki, M., Whittaker, J., Hingorani, A.D., Day, I.N. & Humphries, S.E. (2011) Genetic variants associated with Von Willebrand factor levels in healthy men and women identified using the HumanCVD BeadChip. *Ann Hum Genet,* **75,** 456-467.

Zandberg, W.F., Kumarasamy, J., Pinto, B.M. & Vocadlo, D.J. (2012) Metabolic Inhibition of Sialyl-LewisX Biosynthesis by 5-Thiofucose Remodels the Cell Surface and Impairs Selectin-Mediated Cell Adhesion. *J Biol Chem*.

Zhang, B. & Zhang, Y. (2009) Mann-Whitney U test and Kruskal-Wallis test should be used for comparisons of differences in medians, not means: comment on the article by van der Helm-van Mil et al. *Arthritis Rheum,* **60,** 1565; author reply 1565.

Zhang, Z.P., Blomback, M., Egberg, N., Falk, G. & Anvret, M. (1994) Characterization of the von Willebrand factor gene (VWF) in von Willebrand disease type III patients from 24 families of Swedish and Finnish origin. *Genomics,* **21,** 188-193.

Zhou, Y.F., Eng, E.T., Zhu, J., Lu, C., Walz, T. & Springer, T.A. (2012) Sequence and structure relationships within von Willebrand factor. *Blood,* **120,** 449-458.

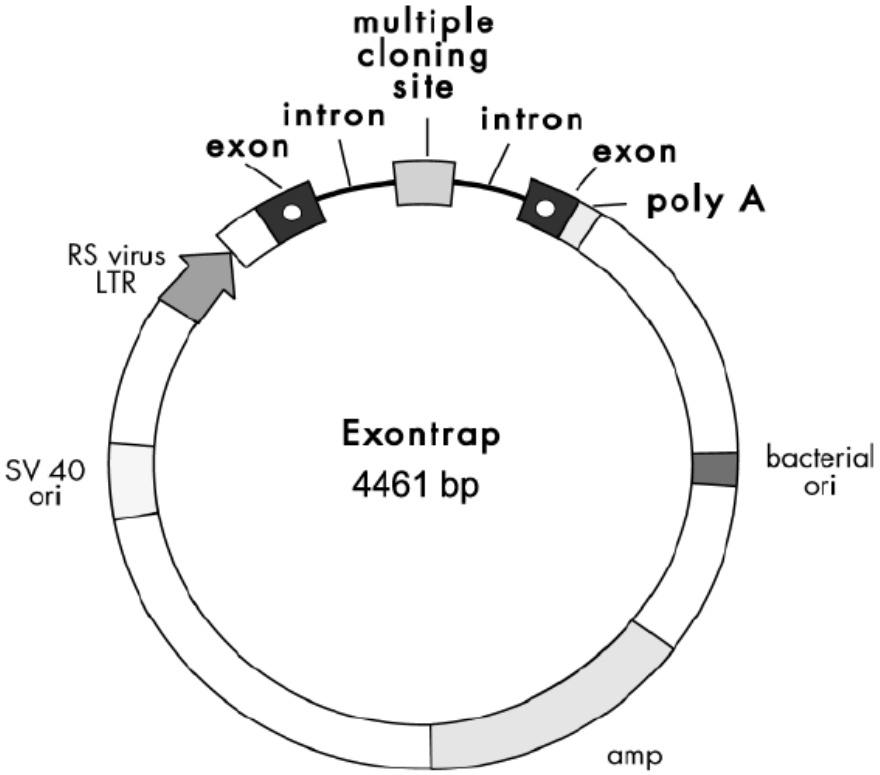
Zhou, Z., Yu, F., Buchanan, A., Fu, Y., Campos, M., Wu, K.K., Chambless, L.E., Folsom, A.R., Boerwinkle, E. & Dong, J.F. (2014) Possible race and gender divergence in association of genetic variations with plasma von Willebrand factor: a study of ARIC and 1000 genome cohorts. *PLoS One,* **9,** e84810.

9 Appendices

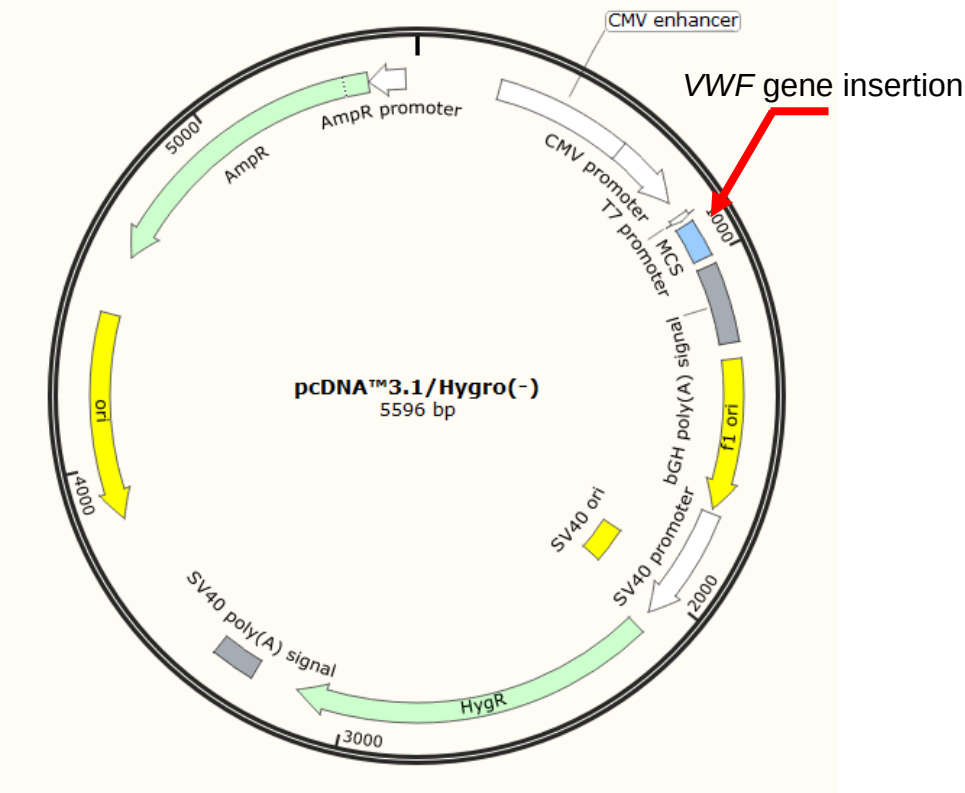
Appendix 1: VWF SNV analysed in this study

RS2362481	RS216867	RS60831677	RS763580
RS12580835	RS216872	RS33978901	RS2238103
RS7310736	RS2058473	RS216321	RS2238101
RS7314566	RS216883	RS216322	RS1034933
RS11063951	RS216893	RS216326	RS2239140
RS1990326	RS216896	RS216327	RS11064024
RS933408	RS12579603	RS216332	RS12317079
RS2286646	RS2229446	RS1860365	RS7306706
RS7962217	RS34230288	RS216336	RS3782716
RS7969672	RS216902	RS10849378	RS12306928
RS11063953	RS216903	RS216339	RS2286608
RS10849362	RS216904	RS2239162	RS10849387
RS10849363	RS216905	RS1063857	RS10774394
RS2363309	RS216801	RS1063856	RS6489695
RS12369177	RS216809	RS216293	RS7979747
RS11063961	RS216811	RS16932374	RS10774401
RS4764478	RS216812	RS7954351	RS10849406
RS35335161	RS542993	RS7312411	RS7956209
RS2270239	RS1800385	RS980131	RS10744699
RS11063965	RS11063987	RS4764482	RS11064074
RS917857	RS216312	RS1800378	RS10774407
RS917858	RS16933969	RS1800376	RS1558327

RS4764521	RS56068059	RS1800375	RS1558325
RS216855	RS11611917	RS11837584	RS887475
RS216856	RS11612384	RS2239153	RS2363880
RS216865	RS1800380	RS2238104	
RS758730	RS797775	RS797773	



Appendix 2: pET01 plasmid used for acceptor / donor motif splicing assay (MoBiTec molecular biotechnology).



Appendix 3: pcDNA3.1/Hygro (-) plasmid map with *VWF* insertion (Figure generated using SnapGene viewer v.2.1).