



The
University
Of
Sheffield.

Copy number variation in von Willebrand disease: Screening, prevalence and functional characterisation

Simon James Webster BSc (Hons)

Thesis submitted for the degree of Doctor of Philosophy

The University of Sheffield

Department of Infection, Immunity and Cardiovascular Disease

Sheffield
October 2016

ABSTRACT

von Willebrand disease (VWD) is the most common mucocutaneous bleeding disorder in humans. It is caused by genetic defects/disruptions in the von Willebrand factor gene (*VWF*) locus. *VWF* encodes a large multimeric glycoprotein (VWF), essential for platelet dependent primary haemostasis and binding, and transportation of coagulation factor VIII. VWF is stored in specialised storage organelles called Weibel-Palade bodies (WPB).

Disruption of *VWF* may result in qualitatively defective (type 2), or quantitatively deficient (types 1 and 3) VWD. It has been reported that in type 1 VWD, ~35% of patients have no causative *VWF* mutation. Copy number variation (CNV) consisting of large exonic deletions and/or duplications within *VWF* have been reported. However, these analyses were limited to exons. The initial aim of this thesis was to design a custom *VWF* microarray to enable comprehensive array comparative genomic hybridisation (aCGH) based CNV screening across the entire *VWF* locus including exonic, intronic and 5' and 3' flanking intergenic regions.

Custom aCGH CNV analyses identified a range of exonic CNV in both type 1 and type 3 VWD patients. In addition two novel 5' intergenic deletions were identified ~9 Kb and 22 Kb upstream of the ATG initiator site. *In silico* analysis of these deletions suggested that they possess active regulatory elements and may be involved in the regulation of *VWF* gene expression. A multiplex genotyping assay was developed to enable high-throughput screening of these deletions in the population.

The pathogenic contribution of *VWF* exonic CNV to quantitative VWD is poorly understood. In this study *in vitro* analysis of three *VWF* exonic deletions, involving exons 3, 32-34 and 33-34 demonstrated defects in VWF secretion and biosynthesis. These findings were supported by high resolution, widefield fluorescent, and super resolution, structured illumination microscopy (SIM). These data revealed profound defects on VWF storage and WPB biogenesis.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the help of a number of important people. Firstly, I would like to say a huge thank you to my primary supervisor Professor Anne Goodeve and co-supervisor Dr. Daniel Hampshire, for taking me on as a PhD student and their continued guidance, patience and unwavering optimism over the past three years. I would also like to thank Professor Ian Peake for all his advice and entertaining and insightful discussions.

It has been a privilege and a pleasure to be part of the Haemostasis Research Group and I have been lucky enough to work with some great colleagues. Special mention must go to past members Dr Vincenzo Leo, Dr Ahmad Mufti, Dr Essa Sabi and Dr Ashley Cartwright who all welcomed me to the group and helped me feel at home from day one. I would also like to thank Professor Martina Daly for offering me the chance to continue as a Postdoc in the group and for having faith that I would be written up on time.

The microscopy performed as part of these studies would not have been possible without the help of Dr Darren Robinson and Dr Christa Walther. I am immensely grateful for their patience, expertise and advice.

I am indebted to my parents who have supported and encouraged me over the years and I am sure they are delighted to have completed the set of PhD children. My brothers Steve and Chris have been an invaluable source of scientific knowhow and discussion, for which I am truly grateful.

Finally, a special thank you to my wife Natalie. Through no fault of your own you have been on this journey with me through all the highs and the lows. Your love, patience and understanding have kept me going and made this experience so much more manageable. We can now look forward to the next chapter in our lives together and for that I cannot wait.

CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
Publications and presentations from this thesis	x
List of Amino Acids	xi
List of Abbreviations	xii
List of Figures	xvi
List of Tables	xix
1 Introduction	1
1.1 Historical Aspects	2
1.2 Von Willebrand Factor (<i>VWF</i>)	3
1.2.1 The <i>VWF</i> Gene	3
1.2.2 <i>VWF</i> Structure	3
1.2.3 <i>VWF</i> biosynthesis, storage and secretion	5
1.3 Haemostasis and the role of <i>VWF</i>	9
1.3.1 Sub-endothelial Matrix Protein Adhesion and Platelet binding	10
1.3.2 FVIII Binding	12
1.4 Proteolytic processing via ADAMTS13	12
1.5 von Willebrand disease (<i>VWD</i>)	13
1.5.1 Type 1 <i>VWD</i>	14
1.5.2 Type 2 <i>VWD</i>	15
1.5.3 Type 3 <i>VWD</i>	15
1.6 Type 1 Disease Mechanisms: <i>in vitro</i> and <i>in vivo</i> studies	16
1.6.1 Decreased <i>VWF</i> synthesis	16
1.6.2 Intracellular retention and defective storage	17
1.6.3 Accelerated clearance	18
1.7 The ABO blood group effect	19
1.8 Other genetic loci associated with <i>VWF</i> levels	20
1.9 Ethnicity effect	22
1.10 Hypotheses and research aims	22
1.10.1 Hypotheses	23
1.10.2 Aims	24
2 Materials and Methods	25
2.1 Materials	26
2.1.1 List of stock solutions	26
2.1.2 Microbiology reagents	27

2.1.2.1	Plasmids and VWF expression vectors.....	27
2.1.2.2	Bacterial culture media.....	27
2.1.2.3	Plasmid purification reagents and kits.....	28
2.1.2.4	TA cloning reagents	28
2.1.2.5	Mutagenesis reagents.....	28
2.1.3	General molecular biology reagents	28
2.1.3.1	PCR and agarose gel electrophoresis.....	28
2.1.3.2	PCR, genotyping, mutagenesis and sequencing primers	29
2.1.4	Array Comparative Genomic Hybridisation (aCGH) reagents	29
2.1.5	Mammalian cell culture and transfection.....	32
2.1.5.1	Human Umbilical Vein Endothelial Cell (HUVEC) culture	32
2.1.5.2	Human Embryonic Kidney (HEK) 293 Cell culture	32
2.1.5.3	HEK 293 cell freezing media reagents.....	32
2.1.5.4	Transfection reagents.....	32
2.1.5.5	Enzyme Linked Immunosorbent Assay (ELISA) and related reagents	32
2.1.6	Immunocytochemistry and fluorescent microscopy reagents	33
2.1.6.1	Antibodies.....	33
2.1.6.2	Non-immunological stains	33
2.1.6.3	Mounting media.....	34
2.1.7	Patient cohort DNA samples.....	34
2.1.8	Laboratory equipment.....	35
2.1.9	Image Analysis.....	36
2.2	Methods	36
2.2.1	Array Comparative Genomic Hybridisation (aCGH) introduction.....	36
2.2.1.1	DNA quantification and quality analysis for use in aCGH	36
2.2.1.2	Sodium acetate and ethanol precipitation DNA clean-up.....	38
2.2.1.3	DNA integrity analysis	38
2.2.1.4	Custom <i>VWF</i> microarray design	39
2.2.1.5	Array design iterations and modifications.....	39
2.2.1.6	Fluorescent labelling of DNA for aCGH.....	40
2.2.1.7	Clean-up of fluorescently labelled DNAs.....	40
2.2.1.8	Hybridisation.....	41
2.2.1.9	Array washing, scanning and feature extraction	41
2.2.1.10	Feature extraction, data analysis and aberration detection.....	41
2.2.2	Polymerase Chain Reaction (PCR)	42
2.2.2.1	Primer design	42
2.2.2.2	Standard PCR	42

2.2.2.3	OneTaq Hot-Start PCR	42
2.2.2.4	Breakpoint mapping	43
2.2.2.5	Expand long template PCR protocol	43
2.2.2.6	Multiplex PCR.....	44
2.2.2.7	Multiplex genotyping.....	45
2.2.2.8	DNA sequencing and sequence analysis.....	45
2.2.3	General molecular biology methods	45
2.2.3.1	Agarose gel electrophoresis of DNA	45
2.2.3.2	QIAquick gel extraction	46
2.2.3.3	TA cloning	46
2.2.3.4	Mutagenesis.....	47
2.2.3.5	Transformation of competent cells	48
2.2.3.6	Growth and storage of E.coli used for plasmid preparations.....	48
2.2.3.7	Plasmid DNA purification (Miniprep)	48
2.2.3.8	Plasmid DNA purification (Maxiprep)	49
2.2.3.9	Quantification of plasmid DNA	50
2.2.4	Mammalian cell culture and <i>in vitro</i> expression	51
2.2.4.1	HEK293 cell culture.....	51
2.2.4.2	HUVEC cell culture.....	51
2.2.4.3	Freezing and long-term storage of HEK293 cells.....	51
2.2.4.4	Cell counting and seeding.....	52
2.2.4.5	Transfections.....	52
2.2.4.6	Cell lysis	53
2.2.5	ELISA.....	53
2.2.5.1	ELISA data analysis	57
2.2.6	Immunocytochemistry and fluorescent microscopy	57
2.2.6.1	Immunocytochemistry	57
2.2.6.2	Widefield fluorescent microscopy.....	58
2.2.6.3	Structured Illumination Microscopy (SIM) and 3D visualisation	58
2.2.7	Statistical analysis.....	59
3	Design, validation and application of a custom microarray for the detection of CNV in VWF	60
3.1	Introduction	61
3.1.1	The evolutionary origin of CNV. Selective advantages and medical relevance.....	62
3.1.2	CNV in VWD	62
3.1.3	Detecting CNV	66
3.1.3.1	MLPA.....	66

3.1.3.2	Array comparative genomic hybridisation	67
3.2	Hypotheses	69
3.2.1	Aims	69
3.3	Results	70
3.3.1	Patient DNA quality assay	70
3.3.2	Custom VWF array design: version 1	70
3.3.3	Array Validation	73
3.3.3.1	Initial validation analysis of design version 1	73
3.3.3.2	Detailed manual probe calling is necessary for pseudogene CNV detection in design version 1	82
3.3.3.3	Detection of a novel 5' intergenic region (IR) deletion in two IC.....	83
3.3.4	Design iteration 2: 8x15k v2.....	86
3.3.5	Design iteration 3: 8x60k	88
3.3.6	Array based CNV screening in a type 1 VWD patient cohort.....	91
3.3.6.1	Heterozygous deletion of exons 4-5.....	91
3.3.6.2	Heterozygous deletion of exon 19-20.....	93
3.3.6.3	Heterozygous deletion of exon 33-34.....	93
3.3.6.4	Heterozygous duplication of exons 23-34	96
3.3.7	Array based CNV screening in type 2 VWD patients.....	96
3.3.8	Array based CNV screening in type 3 VWD patients.....	97
3.3.8.1	3WINTERS type 3 cohort	97
3.3.8.2	Canadian type 3 cohort	100
3.3.9	Detection and interpretation of two novel deletions in the <i>VWF</i> 5' intergenic region (IR).....	102
3.3.10	Incidental findings	105
3.4	Discussion.....	105
4	Breakpoint mapping, <i>in silico</i> characterisation and screening of 5' intergenic deletions found at the <i>VWF</i> locus.....	114
4.1	Introduction	115
4.1.1	Non-coding DNA and ENCODE.....	115
4.1.2	Chromatin, nucleosomes and histones.....	116
4.1.3	DNase-I hypersensitivity sites.....	116
4.1.4	Enhancers, insulators and silencers	117
4.1.5	Non-coding CNV in disease.....	119
4.1.6	<i>VWF</i> transcriptional regulation.....	120
4.2	Rationale and hypotheses.....	122
4.2.1	Aims	123
4.3	Results	123

4.3.1	5'del-I and del-II breakpoint mapping.....	123
4.3.2	<i>In silico</i> analysis of breakpoint architecture and determination of mechanism for 5'del-I.....	127
4.3.3	<i>In silico</i> analysis of breakpoint architecture and determination of mechanism for 5'del-II.....	129
4.3.4	<i>In silico</i> analysis of the VWF 5' intergenic landscape	131
4.3.5	<i>In silico</i> analysis of regulatory activity in the 5'del-I region	131
4.3.6	<i>In silico</i> analysis of regulatory activity in the 5'del-II region	135
4.4.1	Multiplex PCR optimisation	139
4.4.2	5' del genotyping assays.....	141
4.4.3	Prevalence of 5' intergenic deletions in the healthy control cohort.....	141
4.4.4	Correlation of 5' deletions with VWF:Ag levels	143
4.5	Discussion.....	146
5	<i>In vitro</i> functional characterisation of VWF exonic deletion copy number variants causing type 1 VWD	152
5.1	Introduction	153
5.1.1	Aims.....	158
5.2	Results	159
5.2.1	Optimisation of pseudo-WPB production in HEK293 cells.....	159
5.2.2	Generation of rVWF32-34del in pCI-neo	163
5.2.3	<i>In vitro</i> characterisation: ELISA analysis of VWF:Ag	164
5.2.3.1	Reduced biosynthesis as a disease mechanism.....	164
5.2.3.2	Reduced secretion as a disease mechanism.....	167
5.2.3.3	Effective secretion is a product of reduced total VWF and defective secretion	169
5.2.4	Widefield fluorescent imaging of VWF intracellular location and storage	171
5.2.5	Structured illumination microscopy of intracellular VWF and WPB.....	175
5.2.5.1	SIM imaging of pseudo-WPB formation in HEK293 cells transfected with deletion mutant VWF.....	177
5.2.6	Disruption of the signal peptide cleavage site as a potential mechanism for ER retention in VWF exon 3 deletion.....	179
5.3	Discussion.....	181
6	Final discussion and future work	191
6.1	Summary.....	192
6.2	Limitations of research.....	197
6.3	Future work	198
6.3.1	Mutation screening in 5' intergenic region	198
6.3.2	ER stress and Regulated Ire1a Dependent Decay (RIDD) in VWD.....	198

6.3.3	ER stress and autophagy in VWD	203
6.4	Concluding remarks	204
7	References	205
8	Appendices	239
8.1	Appendix 1	240
8.2	Appendix 2	241
8.3	Appendix 3	244
8.4	Appendix 4	246
8.5	Appendix 5	246

Publications and presentations from this thesis

Webster SJ, Hampshire DJ, Theophilus B, Schneppenheim R, Bellissimo D, Peake IR and Goodeve AC. Development and validation of array comparative genomic hybridisation (aCGH) in VWD genetic analysis. Oral presentation at the 2014 British Society for Haemostasis and Thrombosis Annual Meeting, Edinburgh, UK.

Webster SJ, Hampshire DJ, Schneppenheim R, Bellissimo D, James P, Theophilus B, Peake IR, Goodeve AC and on behalf of on behalf of the EU-VWD and ZPMCB-VWD study groups (2015). Detection of large exonic and intergenic deletions in the *VWF* locus of VWD patients using array comparative genomic hybridisation (aCGH). *J Thromb Haemost* **13 (Suppl. 2)**, abst. OR086. (Oral presentation at the International Society for Thrombosis and Haemostasis Scientific Meeting in Toronto, June 2015).

Webster SJ, Cartwright A, Hampshire DJ, Peake IR, Goodeve AC and on behalf of the EU-VWD and ZPMCB-VWD study groups (2015). High resolution microscopic characterisation of VWF biosynthesis, storage and secretion in type 1 VWD patients with large in-frame VWF deletions. *J Thromb Haemost* **13 (Suppl. 2)**, abst. AS064. (Oral presentation at the International Society for Thrombosis and Haemostasis Scientific Meeting in Toronto, June 2015). **This abstract received a Young Investigator Award from the ISTH.**

Webster SJ, Cartwright A, Hampshire DJ, Peake IR, Goodeve AC and on behalf of the EU-VWD and ZPMCB-VWD study groups. High resolution microscopic characterisation of VWF biosynthesis, storage and secretion in type 1 VWD patients with large in-frame VWF deletions. Oral presentation at the 2015 British Society for Haemostasis and Thrombosis Annual Meeting, London, UK.

List of Amino Acids

Amino acid	Single letter code	Three letter code
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamine	Q	Gln
Glutamic acid	E	Glu
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

List of Abbreviations

aCGH	Array comparative genomic hybridisation
ADAMTS13	A disintegrin and metalloproteinase with a thrombospondin type 1 motif 13
ADM	Aberration detection method
AFM	Affected Family Member
Alu	Repetitive element characterised by action of <i>Arthrobacter luteus</i> restriction endonuclease
ATG7	Autophagy related gene 7
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
cDNA	Complementary deoxyribonucleic acid
CK	Cysteine Knot
CLEC4M	C-type lectin domain family 4 member M protein
CNV	Copy number variation
Cy3-dUTP	Cyanine 3-deoxyuridine triphosphate
Cy5-dUTP	Cyanine 5-deoxyuridine triphosphate
DAPI	4',6-diamidino-2-phenylindole
dH₂O	Deionised water
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediamine tetra acetic acid

ELISA	Enzyme linked immunosorbent assay
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated degradation
EU	European Union
FAM	Carboxyfluorescein
FBS	Foetal bovine serum
FVII	Coagulation factor VII
FVIII	Coagulation factor VIII
FVIII:C	Coagulation factor VIII activity
g	G-force
GpI$\beta$$\alpha$	Platelet receptor Glycoprotein I β α
GPIIβ/IIIα	Platelet receptor Glycoprotein II β /III α
GWAS	Genome Wide Association Study
HC	Healthy Control
HEK293	Human embryonic kidney cell 293
HRP	Horse radish peroxidase
HMW	High molecular weight
HUVEC	Human umbilical vein endothelial cell
IC	Index Case
IU/DI	International units per decilitre
IR	Intergenic region
Kb	kilo base
kDa	kilo Daltons
LB	Luria Bertani

LINE	Long Interspersed Nuclear Element
LR-PCR	Long-range polymerase chain reaction
MCMDM-1VWD	Molecular and Clinical Markers for the Diagnosis and Management of type 1 von Willebrand Disease
Mb	Megabase
MgCl₂	Magnesium Chloride
mg	Milligram
µg	Microgram
ml	Millilitre
µl	Microliter
MLPA	Multiplex ligation-dependent probe Amplification
MMEJ	Microhomology mediated end joining
µM	Micro molar
ng	Nanogram
nm	Nanometer
NHEJ	Non homologous end joining
OPD	o-Phenylenediamine
PBS	Phosphate buffered saline
PDI	Protein Disulphide Isomerase
pmol	pico mole
PCR	Polymerase chain reaction
RIPA	Ristocetin induced platelet aggregation
SCARA5	Scavenger receptor class A 5
SIM	Structure illumination microscopy

SINE	Short interspersed nuclear element
siRNA	Small interfering ribonucleic acid
SNP	Single nucleotide polymorphism
STAB2	Stabilin-2 gene
STX2	Syntaxin-2
STXBP5	Syntaxin-binding protein 5 gene
Taq	<i>Thermus aquaticus</i>
TBE	Tris-borate-EDTA
TE	Tris-EDTA
TGN	trans-Golgi network
T_m	Melting temperature
UFM	Unaffected family member
UTR	Untranslated region
V2R	Vasopressin 2 receptor gene
v/v	Volume/Volume
VWD	von Willebrand disease
VWF	von Willebrand factor gene
VWF	von Willebrand factor
VWF: Ag	VWF Antigen
VWF: CB	VWF: Collagen Binding
VWF: RCo	VWF: Ristocetin cofactor activity
WPB	Weibel-Palade bodies
WT	Wild-type
w/v	Weight/Volume

List of Figures

Figure 1.1	Modular domain structure and function of VWF	4
Figure 1.2	Overview of VWF biosynthesis, WPB biogenesis and secretion	8
Figure 1.3	Role of VWF in haemostasis	11
Figure 2.1	Overview of aCGH	37
Figure 2.2	Cell counting using a haemocytometer	55
Figure 2.3	ELISA overview	56
Figure 3.1	Example of DNAOK! quality assay results	71
Figure 3.2	Probe coverage across VWF in custom array design	72
Figure 3.3	IC UK-M C48, heterozygous whole gene deletion of VWF	75
Figure 3.4	IC UK D17, homozygous whole gene deletion of VWF	75
Figure 3.5	IC EU P9 F11 I1, 1.6kb deletion of exon 3	79
Figure 3.6	IC LM12567, duplication of exons 9 and 10	79
Figure 3.7	IC 1585PM pseudogene region deletion (exons 33-34)	84
Figure 3.8	IC P9F3II1 pseudogene region deletion (exons 32-34)	84
Figure 3.9	Example array profiles of the same IC, highlighting the differences between the 3 array designs	89
Figure 3.10	Detection of two heterozygous deletions of exon 4-5	92
Figure 3.11	Detection of a heterozygous deletion of exons 19-20	94
Figure 3.12	Detection of a heterozygous deletion of exons 33-34	95
Figure 3.13	Detection of a pseudogene region duplication	96
Figure 3.14	Detection whole gene deletions in type 3 patients from the 3WINTERS study cohort	99
Figure 3.15	Detection of multi-exonic and intergenic heterozygous deletions in type 3 patients from the 3 WINTERS study cohort	101
Figure 3.16	Detection of 5'del-I and 5'del-II in P9F3	103

Figure 3.17	Two patients from the type 1 VWD cohort, where the only CNV identified was 5'del-I	104
Figure 3.18	Incidental CNV findings in two patients	106
Figure 3.19	Overview of CNVs identified via aCGH in this study	110
Figure 4.1A	PCR across 5'del-I region	126
Figure 4.1B	PCR across 5'del-II region	126
Figure 4.2	MMEJ deletion mechanism of 5'del-I	128
Figure 4.3	Alu-mediated NAHR mechanism in the formation of 5'del-II	130
Figure 4.4	5' Intergenic regulatory landscape of <i>VWF</i>	133
Figure 4.5	Detailed view of 5'del-I region	134
Figure 4.6	Detailed view of 5'del-II region	136
Figure 4.7	5'del multiplex PCR design	138
Figure 4.8	Optimisation of 5'del multiplex PCR and analysis of different 5'del genotypes	140
Figure 4.9	Fluorescent genotyping data for 5' deletions	142
Figure 4.10	Correlation of 5' deletions with VWF:Ag levels in all blood group individuals	144
Figure 4.11	Correlation of 5' deletions with VWF:Ag levels in non-O blood group individuals	145
Figure 4.12	Correlation of 5' deletions with VWF:Ag levels in blood group O individuals	145
Figure 5.1	Optimisation of pseudo-WPB formation in HEK293 cells	160
Figure 5.2	Sequencing confirmation of rVWF32-34del in pCI-neo-VWF	163
Figure 5.3	Quantification of total VWF:Ag in the deletion mutants as a percentage of WT	166
Figure 5.4	VWF secretion ratios for the deletion mutants normalised to WT	168
Figure 5.5	Effective secretion of VWF in WT and deletion mutants	170

Figure 5.6	Widefield fluorescent microscopy of homozygous deletions	173
Figure 5.7	Widefield fluorescent microscopy of heterozygous deletions	174
Figure 5.8	Structured illumination microscopic comparison of WPB formation and morphology in HEK293 cells and HUVECs	176
Figure 5.9	SIM imaging of pseudo WPB formation in WT and deletion mutants	178
Figure 5.10	<i>In silico</i> analysis of the VWF signal peptide disruption in del Ex3	180
Figure 6.1	VWF mRNA sequence showing IRE1 cleavage motifs	200
Figure 6.2	VWF mRNA secondary structure and stem loop formation	201
Figure 6.3	Overview of proposed ER stress effect on VWF biosynthesis	202

List of Tables

Table 1.1	Summary of VWD types	13
Table 2.1	VWF 5' intergenic region deletion primers	30
Table 2.2	Mutagenesis primers used to generate rVWF32-34del	30
Table 2.3	Internal sequencing primers for VWF cDNA sequencing	31
Table 2.4	Thermal cycling parameters for Expand Long Template PCR	44
Table 3.1	Summary of CNV reported in <i>VWF</i>	64
Table 3.2	List of index cases with previously identified CNV across the <i>VWF</i> locus used in the validation study	74
Table 3.3	Comparison of mapped breakpoints to aCGH breakpoints	78
Table 3.4	IC with no breakpoint data	81
Table 3.5	IC with CNV in the pseudogene region	85
Table 3.6	Comparison of validation samples analysed on the 3 array designs	90
Table 3.7	Type 3 VWD patients analysed in this study	98
Table 4.1	Markers of regulatory potential identified in 5'del-I	135
Table 4.2	Markers of regulatory potential identified in 5'del-II	137
Table 4.3	Frequency of VWF 5' deletions in this study and 1000 Genomes	143
Table 5.1	Phenotypic data of IC and AFM with exonic deletion CNV	156

1 Introduction

1.1 Historical Aspects

In 1926 Dr. Erik von Willebrand described a severe mucocutaneous bleeding problem in a large family living on the Åland Islands archipelago in the Baltic Sea. Distinct from haemophilia, the disease was autosomally inherited and characterized by a prolonged bleeding time. von Willebrand termed the disorder “hereditary pseudohemophilia” (von Willebrand, 1999).

This initial report, published in the Swedish Finska Läkaresällskapetets Handlingar Journal in 1926, was the start of an intense research effort into the pathogenesis of this disease that has continued for almost a century.

Early studies led to the opinion that some distinct plasma factor was missing or defective in patients with this disorder. Patients were shown to have decreased FVIII levels and prolonged coagulation time. Importantly, this could be corrected via the transfusion of normal plasma (Sadler, 1998). Transfusions with plasma fraction I-0 led to a reduction in bleeding time and an increase in FVIII coagulant activity (FVIII:C) over 24h. Furthermore, transfusions with fraction I-0 from haemophilia A patients also corrected the VWD symptoms, highlighting a clear distinction between the two diseases (Nilsson et al., 1957).

It was not until 1971, some forty-five years after von Willebrand’s initial observation, that this factor was immunologically detected and differentiated from FVIII (Zimmerman et al., 1971). Using anti-serum generated against FVIII, it was shown that factor VIII-related antigen (FVIII_{Rag}) was reduced in VWD patients but not in patients with haemophilia or healthy controls. Today, FVIII_{Rag} is known as von Willebrand factor (VWF).

Advances in molecular biology and genetics in the years since its discovery have shed new light on VWF structure, function and dysfunction in VWD.

1.2 Von Willebrand Factor (VWF)

1.2.1 The VWF Gene

Following the detection of VWF, a number of groups simultaneously began to decipher its coding sequence using cDNA libraries (Ginsburg et al., 1985, Lynch et al., 1985, Sadler et al., 1985, Verweij et al., 1985). The VWF gene (*VWF*) is located on chromosome 12p13 (Verweij et al., 1985), where it is organized into 52 exons spanning 178Kb of genomic DNA (Mancuso et al., 1989). The exons range in size from 40bp up to 1.4Kb (exon 28). The complete pre-pro *VWF* sequence was reported by Bonthron and co-workers in 1986 (Bonthron et al., 1986)(Bonthron et al., 1986). A partial *VWF* pseudogene is located on chromosome 22, replicating *VWF* exons 23-34 with 97% sequence homology (Shelton-Inloes et al., 1987, Mancuso et al., 1989). The combination of partial pseudogene along with the highly polymorphic nature of *VWF*, somewhat complicates the molecular genetic analysis of this locus.

1.2.2 VWF Structure

The initial domain structure of VWF proposed in 1989, describes repeats of four homologous domains in the sequence D1-D2-D'-D3-A1-A2-A3-D4-B1-B2-B3-C1-C2-CK (Mancuso et al., 1989). This has since been superseded by that of Zhou and colleagues (Zhou et al., 2012). This re-annotation utilised a combination of VWF database annotations, knowledge of disulphide bond formation and the homologous domain structure, integrated with sequence analysis and electron microscopy data, in order to present a comprehensive overview of VWF architecture.

This updated structure (Figure 1.1), shows the modular organization of functional segments (VWD, C8, TIL and E) within D assemblies, in addition to the well characterized A domains (Springer, 2011). The B and C domains are now characterised as six tandem von Willebrand-like C domains (VWC).

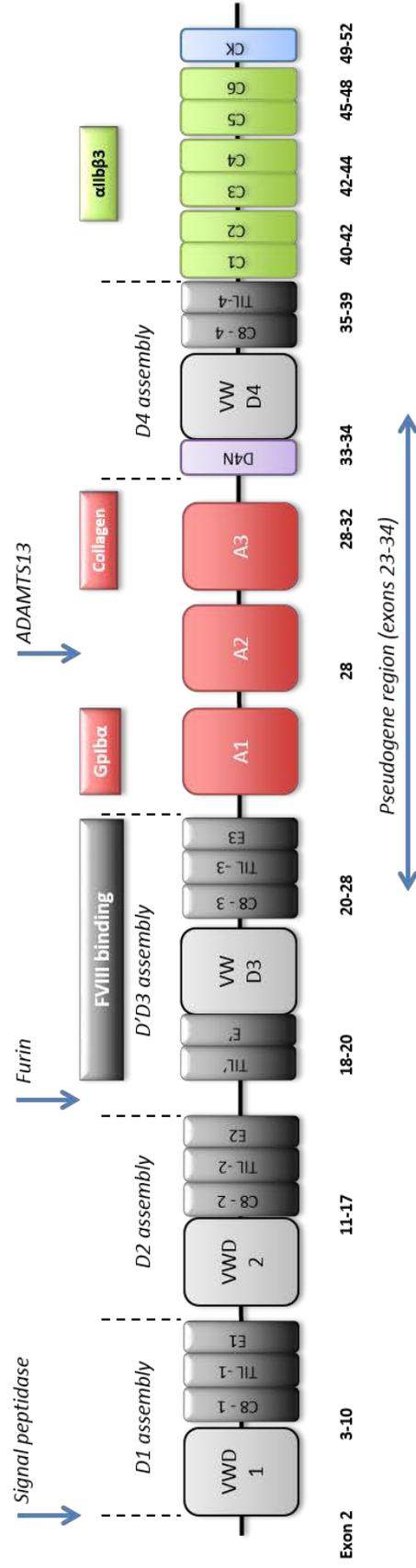


Figure 1.1 Modular domain structure and function of VWF as outlined by Zhou and colleagues (Zhou et al 2012).

VWF is organised into distinct modular assemblies containing the VWD, C8, TIL and E domains. Distinct functional sites exist within these assemblies: D'D3 is the site of FVIII binding, A1 and A3 are involved in binding GpIb α and collagen respectively and the C3 and C4 domains bind the platelet integrin α IIb β 3. Blue vertical arrows indicate cleavage sites for the signal peptide, propeptide (furin site) and ADAMTS13 metalloprotease. The partial VWF pseudogene present on chromosome 22 corresponds to exons 23 – 34.

VWD, VWF D domain; C8, cysteine 8 domain; TIL, trypsin-inhibitor-like domain; E, E repeats; C1-C6, VWF C domains; CK, cysteine knot domain.

1.2.3 VWF biosynthesis, storage and secretion

VWF is a large, multimeric glycoprotein that is synthesized as a 2813 amino acid (aa) preprotein in endothelial cells (EC) and megakaryocytes (Wagner and Marder, 1983, Sporn et al., 1985). This pre-pro VWF consists of a 22 aa signal peptide required for targeting to the endoplasmic reticulum (ER), a 741 aa propeptide and a 2050 aa mature VWF subunit (Sadler, 1998). Within the endoplasmic reticulum intrachain disulphide bond formation occurs within VWF subunits and these VWF monomers dimerize via the protein disulphide isomerase (PDI) initiated formation of disulphide bonds at Cys2771 and 2773 in their C-terminal cysteine knot (CTCK) domains (Katsumi et al., 2000, Lippok et al., 2016). The mature VWF protein is glycosylated and sulfated (Carew et al., 1990), a process that occurs along with propeptide cleavage and multimerisation in the trans-Golgi network (TGN) (Vischer and Wagner, 1994, Wagner et al., 1986). Multimerisation involves the formation of interchain disulphide bonds between neighboring D'D3 modules (Mayadas and Wagner, 1992) (Figure 1.2). In the low pH environment of the trans-Golgi, repeating units formed from the propeptide domains (D1 and D2) and the disulphide linked D'D3 dimer, assemble into the characteristic helical tubules of Weibel-Palade bodies (WPB) (Huang et al., 2008). These specialized secretory organelles, first identified in pulmonary artery endothelial cells (Weibel and Palade, 1964), are the site of VWF multimer synthesis and storage in ECs (Wagner et al., 1982). In megakaryocytes, the VWF multimers are stored within alpha granules.

The Golgi performs a major role in the regulation of WPB biogenesis, size and content. The mini-stack and ribbon like architecture enables the Golgi to limit the size of individual VWF units or “quanta” and package them together into longer organelles (Ferraro et al., 2014). Interestingly, Ferraro *et al* found that reducing VWF levels via siRNA reduced both quantity and length of WPB while the VWF quanta size unit remained the same. The quanta are simply fewer and therefore more dispersed throughout the Golgi, resulting in their individual packaging into shorter WPB.

The developing WPB maintains multiple connections with the Golgi, enabling large quantities of non-tubulated VWF to be added to the forming WPB during its maturation (Mourik et al., 2015a). Consistent with this hypothesis, Mourik and colleagues found that immature WPB contained regions of partially tubulated

VWF in addition to VWF in a non-tubulated state. Both studies from Ferraro et al and Mourik et al, support a model whereby the Golgi may combine numerous small developing WPB to regulate the size of WPB formation.

Following biosynthesis, VWF is secreted from endothelial cells via three distinct routes; constitutively via anterograde secretory carriers (Sporn et al., 1986), basally via the unstimulated, continuous release from WPB (Giblin et al., 2008) and via regulated secretion from WPB following stimulation (Tsai et al., 1991). Recent evidence suggests that each of these secretory routes has a distinct polarity of secretion and that they each give rise to distinct pools of VWF multimers that may have unique physiological significance (Lopes da Silva and Cutler, 2016). This study provided experimental evidence to support the view that ultra-large HMW VWF multimers, released from WPB following stimulation, are secreted apically into the vessel lumen. Furthermore, a distinction between constitutive and basal secretion was demonstrated, indicating that constitutively secreted VWF is via the basolateral direction whereas basal, unstimulated secretion of HMW VWF multimers occurs primarily to the apical side of the cell. This basally secreted VWF includes the full range of VWF multimers including a prominent dimer band and may represent the primary source of plasma VWF.

Unstimulated cells do not readily produce VWF strings upon exocytosis and this fits the hypothesis that basally secreted VWF is the major source of plasma VWF. In the plasma, under normal conditions, prothrombotic VWF string formation is undesirable and could lead to pathogenic thrombotic events. String formation is therefore tightly regulated and localized to the site of vessel damage as part of the regulated secretory pathway. Regulated or stimulated VWF secretion from WPB is induced by agonists that increase intracellular Ca^{2+} or cAMP levels. Stimulated secretion via histamine and thrombin results from an intracellular rise in Ca^{2+} concentration (Hamilton and Sims, 1987, Birch et al., 1992) whereas other agonists such as epinephrine and forskolin cause increases in cAMP and activation of associated signalling pathways. It is thought that the string formation, brought about following stimulated secretion may involve a multigranular model whereby numerous WPB congregate and fuse forming a secretory pod (Valentijn et al., 2010, Mourik et al., 2013). This pooling of WPB contents and mass exocytic

event may promote the formation of strings that is not observed during basal secretion due it consisting of random individual WPB exocytic events. Localised mass exocytosis of VWF and string formation is consistent with the need for a rapid coagulation response following vessel damage.

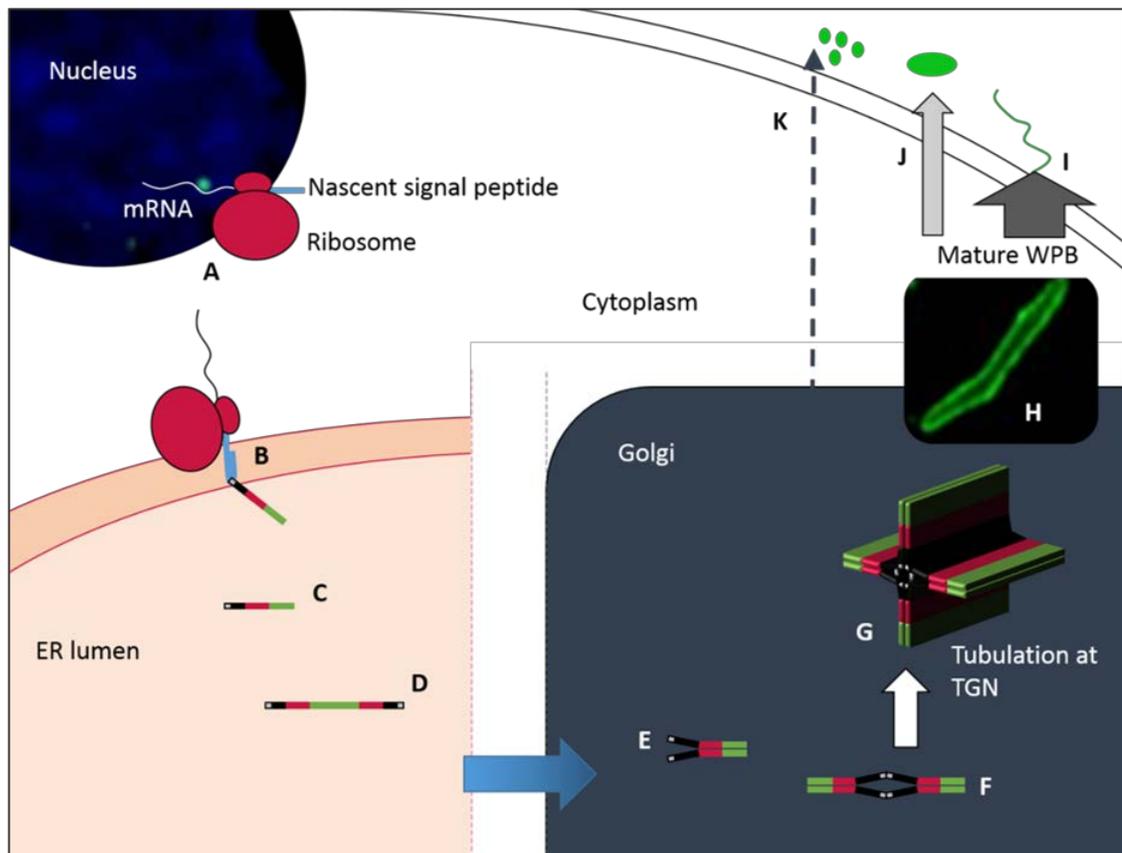


Figure 1.2 Overview of VWF biosynthesis, WPB biogenesis and secretion

A. VWF mRNA ribosome complex initiates translation and emergence of the signal peptide targets the complex to the ER. **B.** The signal peptide anchors the mRNA-ribosomal translation complex to the ER membrane. **C.** cleavage of the signal peptide yields monomeric VWF which undergoes intrachain disulphide bond formation, stabilising the tertiary structure. **D.** VWF dimerization occurs via interchain disulphide bonds between the C-terminal CK domains (tail to tail). **E.** in the low pH of the Golgi, VWF dimers zip-up into a dimeric bouquet. **F.** Multimerisation takes place in the Golgi via interchain disulphide bonds between the N-terminus regions (head to head). **G.** Tubulation occurs at the TGN and furin cleaves the propeptide (which stays non-covalently attached), prior to WPB formation (**H**). Tubulation may also occur in the newly forming WPB. **I.** Secretion from mature WPB may occur via the regulated pathway resulting in the exocytosis of ultra large HMW VWF and string formation. **J.** Basal secretion also occurs from WPB releasing a range of multimeric VWF into the vessel lumen or basolaterally into sub-endothelial matrix. **K.** Constitutive secretion of VWF primarily releases low molecular weight VWF via the basolateral route.

1.3 Haemostasis and the role of VWF

Haemostasis is the complex and regulated coordination of biochemical events that enables the prevention of blood loss from damaged vessels, while under normal conditions maintains blood in a fluid state (Versteeg et al., 2013). It is a vital and therefore highly conserved mechanism from fish to humans and is thought to have evolved over 430 million years ago (Davidson et al., 2003).

Haemostasis can be divided into two principal components: primary haemostasis and secondary haemostasis. Primary haemostasis refers to the binding and aggregation of platelets at the site of vessel injury. This immediate response relies on the combined action of VWF and platelets leading to the formation of a platelet plug. This process is facilitated by VWF through its binding of exposed sub-endothelial collagen. The unfurling of HMW VWF multimers into strings enables the capture and tethering of platelets from the blood flow. This in turn leads to platelet accumulation, activation and stimulation of the coagulation cascade, via the provision of a large procoagulant surface area.

Secondary haemostasis refers to the complex series of interactions of numerous clotting factors that ultimately results in the generation of fibrin. Fibrin stabilises the platelet plug formed during primary haemostasis leading to a stable fibrin clot. Following the identification of many components in this intricate pathway, a cascade model of coagulation was put forward in the 1960's (Davie and Ratnoff, 1964, Macfarlane, 1964). In the classic model, the coagulation cascade consists of two separate pathways, the intrinsic pathway, the components of which are present in the blood and the extrinsic pathway which requires tissue factor (TF), which is exposed on sub-endothelial cells following damage to the vessel. Both pathways converge on the assembly of the prothrombinase complex consisting of activated factor X and factor V (FXa/FVa) which cleaves prothrombin to form thrombin. Thrombin cleaves FVIII from VWF (Hill-Eubanks et al., 1989) and also converts fibrinogen into insoluble fibrin. These pathways however, are not distinct and redundant as both are required for haemostasis *in vivo*. The cell based model of haemostasis suggests that coagulation reactions take place on the cell surfaces of platelets and TF bearing cells, in three distinct stages; initiation, amplification and propagation that link the intrinsic and extrinsic pathways (Monroe and Hoffman, 2006). The initiation phase describes the classic extrinsic

pathway where, TF binds FVII, which is then activated (FVIIa). This TF/FVIIa complex activates small amounts of FIX and FX. This FXa can associate with FVa and form the prothrombin complex, leading to the production of small amounts of thrombin. During the amplification stage, this thrombin can activate platelets that have been recruited to the site of injury and activate FV and FVIII. The propagation phase can then occur on the procoagulant, activated platelet surface, leading to the formation of large amounts thrombin, sufficient to form enough fibrin required for clot formation.

The mosaic structure of the VWF protein is indicative of its multi-functional role and enables it to contribute to numerous aspects of this haemostatic response. Some of these functions in relation to haemostasis are considered below (summarized in Figure 1.3).

1.3.1 Sub-endothelial Matrix Protein Adhesion and Platelet binding

At sites of vascular injury, VWF binds to exposed sub-endothelial collagen via its A3 domain (Romijn et al., 2001), thus forming a “molecular bridge” which tethers platelets, via binding of the A1 domain to the platelet glycoprotein Ib α (GpIb α) (Huizinga et al., 2002). The multimeric size of VWF dictates its platelet binding capacity, as HMW multimers have a higher binding affinity (Federici et al., 1989). Under high shear stress conditions this VWF-dependent platelet adhesion initiates platelet activation and aggregation. (Savage et al., 1996). Platelet aggregation involves the interaction of the platelet integrin α IIb β 3 with the VWC4 module of VWF (Ruggeri et al., 1983, Zhou et al., 2012). Therefore VWF, by tethering platelets to the site of injury initiates and propagates a localized and potent procoagulant response.

This process can also be studied *in vitro*, in the absence of high shear stress, via the use of ristocetin, a glycopeptide antibiotic that induces VWF dependent platelet aggregation (Scott et al., 1991). For this reason, ristocetin has found widespread use as a component in the VWD diagnostic assays: ristocetin cofactor assay (VWF:RCo) and ristocetin-induced platelet aggregation assay (RIPA).

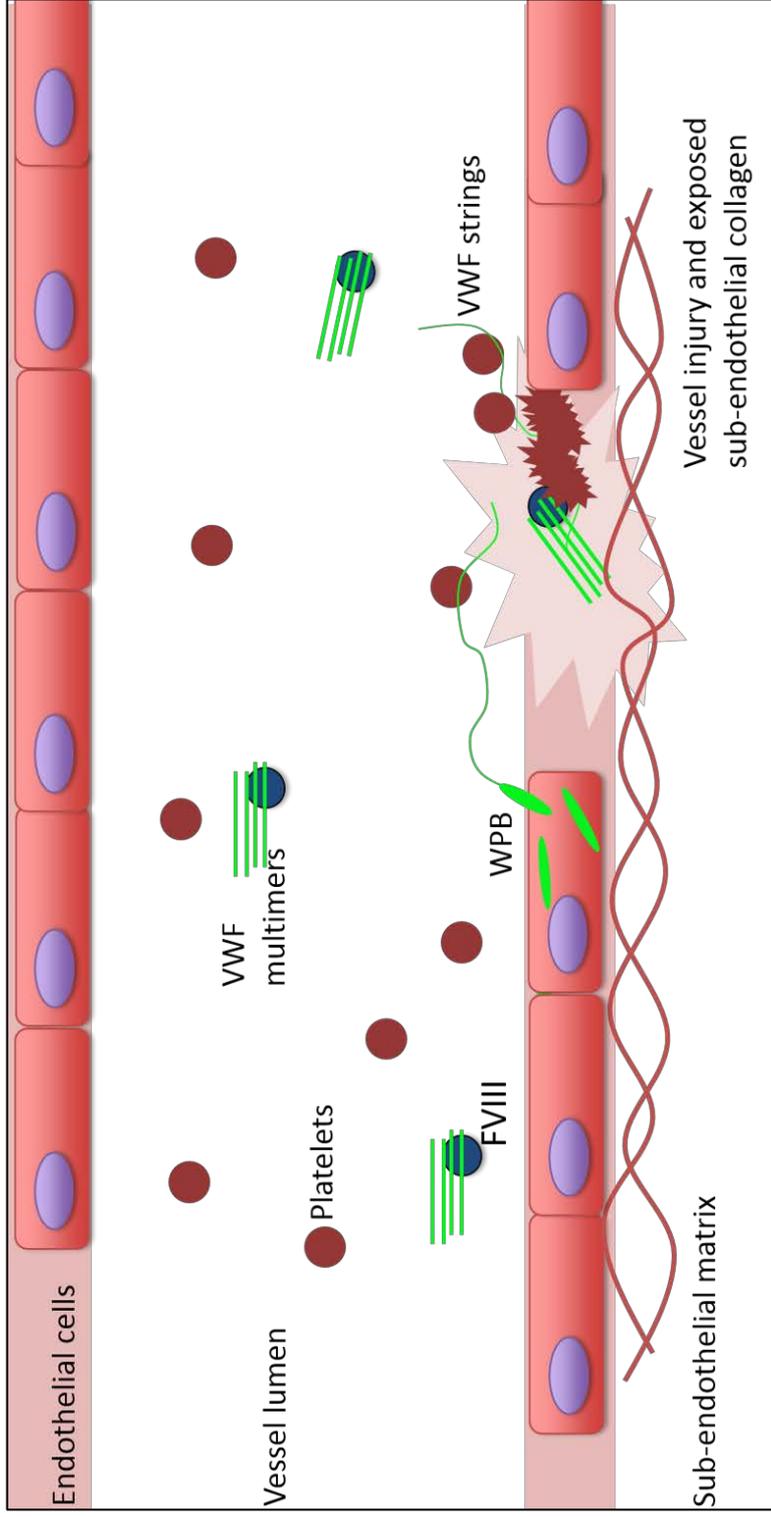


Figure 1.3 Role of VWF in haemostasis. In the circulation VWF binds and stabilizes FVIII. Upon vascular injury, VWF binds the exposed sub-endothelial collagen. Bound VWF is unfurled into string like structures due to shear stress, allowing tethering of platelets to the site of injury, leading to platelet activation and aggregation. FVIII functions in the clotting cascade, which ultimately forms a fibrin clot to stem blood flow from the damaged vessel.

1.3.2 FVIII Binding

Transfusion studies using plasma from both haemophilia A and VWD patients demonstrated the stabilizing activity of VWF on blood coagulation factor VIII (Weiss et al., 1977, Tuddenham et al., 1982) and it is now well established that the survival of FVIII is highly dependent on the formation of VWF-FVIII complexes (Sadler, 1998). Moreover, this interaction ensures functional FVIII is delivered to the site of injury and the platelet plug, where it has a key role in amplification and propagation of the clotting cascade. Binding of FVIII occurs in the amino terminal 272aa residues of the mature peptide (Foster et al., 1987), now known to be in the D'D3 assembly of VWF. High resolution structural studies of D' have also begun to provide insights into the mechanisms of the VWF-FVIII interaction, highlighting the importance of the flexibility and dynamism of the TIL' domain (Shiltagh et al., 2014).

1.4 Proteolytic processing via ADAMTS13

Following its release, VWF undergoes proteolytic processing via the zinc metalloprotease ADAMTS13 (Zheng et al., 2001), first identified via linkage analysis in patients suffering from thrombotic thrombocytopenic purpura (TTP), who were found to have mutations in an ADAMTS candidate gene that mapped to chromosome 9q34 (Levy et al., 2001). Prior work had already purified this VWF protease that cleaved VWF at the p.Tyr1605 – Met1606 peptide bond within the A2 domain (Furlan et al., 1996). Reports of patients with defects in this protease, either through autoimmune mechanisms or mutations demonstrated the role of ADAMTS13 in the generation of different sized VWF multimers. This aspect of VWF biology is a vital regulatory process that prevents pathogenic thrombotic consequences. However, the ADAMTS13 – VWF relationship is also a key factor in accelerated VWF clearance, which along with reduced synthesis and intracellular retention is an identified mechanism of VWD (Badirou et al., 2010).

1.5 von Willebrand disease (VWD)

VWD is the most common autosomally inherited mucocutaneous bleeding disorder, with a prevalence estimated to be up to 1% (Rodeghiero et al., 1987, Werner et al., 1993, Nichols et al., 2008). However, it has been suggested that this may be an over-estimation as clear linkage with the *VWF* locus is not observed in a significant proportion of these cases (Castaman et al., 1999). VWD referrals to tertiary care centres average approximately 1 in 10,000 representing a clear discrepancy in estimations of VWD incidence. A recent epidemiological study aimed at addressing this discrepancy estimated the prevalence of symptomatic VWD to be at least 1 in 1,000 (Bowman et al., 2010).

VWD is caused by mutations within the *VWF* locus, resulting in either quantitative (type 1 and type 3) or qualitative (type 2) defects in VWF. The most recent classification of VWD, provided by the International Society on Thrombosis and Haemostasis Scientific and Standardisation Committee (ISTH-SSC on VWD) describes six distinct types: 1, 2A, 2B, 2M, 2N and 3 (Sadler et al., 2006), summarized in Table 1.1.

Table 1.1 Summary of VWD types

VWD Type	Description of VWF defects
1	Partial quantitative deficiency
2A	Reduced platelet binding associated with loss of HMW multimers
2B	Pathogenic increase in affinity for platelet GpIb α
2M	Reduced platelet binding without loss of HMW multimers
2N	Reduced binding of coagulation factor VIII
3	Virtually complete deficiency of VWF

1.5.1 Type 1 VWD

Type 1 VWD results in a partial quantitative deficiency in VWF and accounts for around 70-80% of all VWD diagnoses in non-consanguineous populations. Diagnosis is complicated by the fact that at the lower end of the normal distribution of VWF levels, healthy individuals are indistinguishable from those with mild type 1 VWD due to *VWF* mutation. Furthermore, type 1 disease is usually a dominant disorder, displaying variability in penetrance and expressivity which can be influenced by multiple environmental and genetic factors (James and Lillicrap, 2013).

In 2006-07 three large, multi-centre studies were published that explored the phenotypic and molecular genetic features of patients historically diagnosed with type 1 VWD. These are referred to as the European (Goodeve et al., 2007), Canadian (James et al., 2007) and UK (Cumming et al., 2006) studies.

These studies analysed 305 type 1 VWD patients, identifying candidate mutations in around 65% of index cases (IC). Missense mutations were the most common in all three studies accounting for 70% of those identified. Other genetic changes included splice site mutations (9%), transcription (8%), small deletions (6%), nonsense (5%) and small insertion or duplication mutations (2%) (Goodeve, 2010). Missense mutations were identified throughout *VWF* and in 15-20% of cases, more than one candidate mutation was identified. In both the European and Canadian studies, the likelihood of finding a candidate mutation increased with reduced VWF levels. The European study also reported that detection of a subtle loss of HMW multimers was observed in ~38% of index cases, 95% of which showed candidate *VWF* mutations.

These studies enabled a number of mutations to be related to type 1 VWD including p.Cys1130Phe and p.Arg1205His, which were observed in patients with low levels of VWF. The most common mutation, detected in 8% of ICs was p.Tyr1584Cys. This variant segregates with *VWF* in a significant proportion of type 1 VWD patients and is thought to be the result of a founder effect (O'Brien et al., 2003). *In vitro* expression studies of this variant have demonstrated increased intracellular retention of VWF and increased susceptibility to ADAMTS13 (O'Brien et al., 2003). Subsequent studies since the publication of the type 1 multi-centre studies have demonstrated that large in-frame exonic

deletions also contribute to the mutation spectrum of type 1 VWD (Sutherland et al., 2009).

1.5.2 Type 2 VWD

Type 2 VWD describes those variants with a qualitative defect in VWF. Currently type 2 VWD is made up of four sub-groups: 2A, 2B, 2M and 2N (Sadler et al., 2006). Type 2A is the most common subtype of type 2 VWD and describes the loss of HMW multimers and therefore a reduced affinity for platelet GpIb α binding and collagen binding. Type 2A is classically caused by mutations in the A2 domain leading to increased proteolysis by ADAMTS-13 (Hassenpflug et al., 2006). Loss of HMW may also be caused by mutations in the CK domain of VWF that disrupt dimerisation, leading to mutant monomers being incorporated during multimerisation (Schneppenheim et al., 1996). In type 2B, missense mutations cause enhanced affinity of GpIb α binding. These gain of function mutations lead to the loss of HMW multimers and may cause thrombocytopenia. Type 2M VWD describes VWF that shows reduced ability to bind GpIb α , often caused by mutations in the A1 domain. In this type there is no loss of HMW multimers as seen in type 2A. 2N may also be caused by mutations in the A3 domain, thus affecting collagen binding. Type 2N VWD is similar in phenotype to mild haemophilia A, due to the defective binding of FVIII by VWF and has a recessive inheritance pattern (Goodeve, 2010).

1.5.3 Type 3 VWD

Type 3 VWD is the rarest form of VWD (~5% of all cases) and also the most severe, resulting as it does in severely reduced or undetectable levels of VWF and FVIII activity (Eikenboom, 2001). Patients with type 3 VWD therefore have defects in primary and secondary haemostasis, resulting in severe bleeding symptoms. Type 3 VWD is classically inherited in a recessive pattern, with affected individuals being either homozygous or compound heterozygous for null alleles. However this concept has been challenged by the reports that some type 3 VWD obligate carriers are symptomatic with increased bleeding symptoms and reduced FVIII activity compared to controls (Castaman et al., 2006, Montgomery, 2006). A number of type 3 cohort studies have been carried out where the

frequently detected mutations include nonsense, small deletions, large deletions, splice site mutations and insertions. Missense mutations are also found in type 3 VWD and likely cause aberrant VWF folding leading retention in the ER (Goodeve, 2010). Partial (single and multiple exons) gene deletions are also reported in type 3 VWD (Peake et al., 1990, Baronciani et al., 2003, Mohl et al., 2008, Sutherland et al., 2009), as are deletions of the entire VWF gene (Schneppenheim et al., 2007).

1.6 Type 1 Disease Mechanisms: *in vitro* and *in vivo* studies

The quantitative deficiency in VWF observed in type 1 VWD occurs via at least three different disease mechanisms: decreased synthesis; intracellular retention; and/or accelerated clearance of functional VWF from the circulation. A number of mutations have been identified that give rise to each of these disease mechanisms. The molecular pathogenesis of these mutations has been primarily elucidated via *in vitro* and in some cases *in vivo* experiments.

1.6.1 Decreased VWF synthesis

Disruption of the VWF promoter and splice sites may give rise to decreased and/or aberrant VWF mRNA. Variation in a short tandem repeat (STR) in the VWF promoter has been shown to modulate VWF levels (Hickson et al., 2011). A deletion within the VWF promoter resulting in aberrant transcription factor binding also leads to a significant reduction in VWF expression (Othman et al., 2010).

The p.R924Q mutation, the second most common variant detected in the Canadian study does not show significantly abnormal VWF synthesis, trafficking or storage. However, mRNA analysis detected a truncated transcript resulting from activation of a cryptic splice site in exon 28 (Berber et al., 2009). In addition, the fact that the contribution of p.Arg924Gln to the type 1 phenotype of patients in the European study is significant, with the c.2771A allele leading to reduced VWF and FVIII levels particularly when in combination with blood group O (Hickson et al., 2010), highlights the difficulty of assigning pathogenicity to a variant based solely on *in vitro* observations. Pruss and co-workers demonstrated reduced biosynthesis of total VWF in the p.Arg1205His and p.Tyr1584Cys

mutations. In this study, *in vitro* results in various cell lines were supported by an *in vivo* mouse model (Pruss et al., 2011). These results supported the observations of these two variants in both HEK293T cells and COS-7 cells and highlighted the utility of the *in vivo* approach to study VWF biosynthesis, clearance and ADAMTS13 cleavage.

The vast majority of VWF expression studies to date have utilised cell lines such as HEK293 cells and COS-7. To overcome some of the inherent limitations of these cell types, primarily that they cannot accurately mirror endothelial cell biology, other approaches have been sought. Starke and colleagues isolated blood outgrowth endothelial cells (BOECs), from peripheral blood of type 1 and type 2 VWD patients (Starke et al., 2013). VWF mRNA levels in three of the four type 1 patient BOECs were decreased, a finding consistent with a defect in synthesis. Interestingly however the p.Tyr1584Cys patient had mRNA levels in the normal range and the defect appeared to relate to abnormal processing of VWF.

1.6.2 Intracellular retention and defective storage

Fourteen of the missense mutations identified in the European study have been expressed *in vitro* (Eikenboom et al., 2009). In COS-7 cells, seven of these variants showed increased intracellular retention and reduced VWF secretion. Homozygous expression of the mutant constructs showed a loss of HMW multimers, which was corrected upon co-transfection with wild-type VWF. Two of these mutations p.Gly160Trp and p.Asn166Ile, lie within the D1 assembly, a region known to be important for disulphide bond formation during multimerisation (Mayadas and Wagner, 1992). Four other variants displayed a mild reduction in VWF secretion and no pathogenicity could be assigned to three of the mutations (p.Gly19Arg, p.Pro2063Ser and p.Arg2313His) using this system. Despite apparently normal function in this expression system, p.Gly19Arg (nucleotide change c.55G>A) may disrupt a donor splice site, leading to aberrant splicing. This hypothesis is supported by *in silico* predictions, and through type 3 VWD in homozygous patients (Schneppenheim, personal communication).

The p.Cys1149Arg mutation in the VWD3 domain causes severe dominant type 1 VWD, and *in vitro*, this variant displayed reduced VWF secretion due to retention in the ER (Eikenboom et al., 1996). Further supporting its role in the dominant type 1 phenotype, co-transfections of WT and mutant VWF showed a dominant suppression of normal VWF secretion in a dosage dependent manner. The mechanism of this observed dominant suppression is now thought to involve the proteosomal degradation of heterodimers consisting of normal and mutant VWF monomers resulting in a 75% reduction in VWF (Bodó et al., 2001). Mutations causing ER retention of misfolded VWF can lead to defective storage in WPB and therefore impaired secretion (Wang et al., 2011). Recent studies using BOECs have demonstrated aberrant morphologies of WPBs for type 1 patients heterozygous for the mutations p.Ser1285Pro, p.Leu1307Pro and p.Cys2693Tyr (Wang et al., 2013a). In these mutants, VWF was retained in the ER and exocytosis of WPBs and formation of VWF strings was reduced.

1.6.3 Accelerated clearance

The reduced survival of VWF is a well-established disease mechanism in type 1 VWD patients (Haberichter et al., 2008). Otherwise known as the Vicenza mutation, p.Arg1205His was previously classified as type 2M and is the classic example of the increased clearance phenotype (Casonato et al., 2002). *In vivo* studies suggest a role for the D'-D3 region in preventing clearance and p.Arg1205His was shown to reduce this protective effect. This mutant also shows reduced affinity for the VWF-propeptide *in vitro* (Lenting et al., 2004). Recent studies support the reduced survival mechanism of Arg1205His and demonstrated a specific effect of this mutation on macrophage-dependent clearance *in vivo* (Rawley et al., 2015).

The amino acid change p.Cys1130Phe also shows increased clearance along with p.Cys1149Arg, p.Cys2671Tyr and p.Trp1144Gly, and this has been demonstrated *in vivo* using a VWF deficient mouse model. Patients with these mutations show increased propeptide/VWF:Ag ratios and increased clearance of VWF from the circulation following desmopressin treatment (Van Schooten et al., 2005).

1.7 The ABO blood group effect

The *ABO* gene encodes a glycosyltransferase, which converts the precursor H antigen into either A or B antigens (N-acetylgalactosamine and D-galactose respectively) depending on the presence of A or B alleles in the individual. In blood group O individuals who lack the transferase activity, only the H antigen is present. The ABO antigens (A, B and H), are known to have a significant association with VWF plasma levels. On VWF, the ABO antigens are attached to Asp-linked oligosaccharide chains (Matsui et al., 1992).

The effect of the ABO blood group on VWF plasma levels is well established with blood group O individuals showing ~30% reduction compared to other blood types (Gill et al., 1987). Furthermore, blood group O individuals show accelerated VWF clearance and a significantly reduced VWF half-life (10h vs. 25h) following treatment with desmopressin (Gallinaro et al., 2008). The O blood group also displays slightly increased susceptibility to ADAMTS13 proteolysis and this has been suggested to be a factor in the accelerated clearance phenotype (Bowen, 2003). However, in ADAMTS13 deficient mice, VWF clearance does not differ significantly from wild-type, indicating a lack of ADAMTS13 involvement and raising the possibility that other proteases may have a role in this process (Badirou et al., 2010).

The ABO effect is most evident in milder forms of VWD as the Canadian study showed that in cases where the VWF level was below 30 IU/dL, blood group O individuals are no longer over represented (James et al., 2007). Of the IC without a candidate mutation recruited in the European study, 76% were blood group O. This increased to 89% in those IC where the disease did not co-segregate with *VWF* (Goodeve et al., 2007). Recent evidence suggests that it may be the blood group status of individuals that influences VWF levels rather than the ABH antigen loading on VWF itself, as no direct effect on clearance was observed in an *in vivo* experimental system designed to monitor A, B and O VWF clearance in the same individual (Groeneveld et al., 2015).

1.8 Other genetic loci associated with VWF levels

Linkage analysis from the European and Canadian studies suggested that the diagnosis of type 1 VWD is linked with the *VWF* gene in ~50% of families (Eikenboom et al., 2006, James et al., 2006), indicating a pathogenic contribution from other genetic loci in addition to *VWF* and also incomplete penetrance.

Both genome wide linkage analysis and genome wide association studies (GWAS) have been used to identify novel genetic loci associated with VWF plasma levels (Souto et al., 2003, Smith et al., 2010), identifying the *ABO* locus as the most significant genetic component on VWF levels. As part of the Cohorts for heart and research in genome epidemiology (CHARGE) consortium, Smith et al (2010) confirmed the association of *ABO* and *VWF* loci with VWF plasma levels and also identified a number of novel candidate loci including *STXBP5*, *SCARA5*, *STAB2*, *STX2*, *TC2N* and *CLEC4M* (Smith et al., 2010).

A number of these loci have since been explored further to elucidate any potential influence they have on VWF levels. The C type lectin member 4 family M (*CLEC4M*) gene was associated with VWF levels and further studies have shown that this lectin receptor can bind and internalize VWF. Variation within *CLEC4M* contributed to variation in VWF plasma levels (Rydz et al., 2013).

More recently the effect of syntaxin-binding protein (*STXBP5*) on VWF levels was investigated and was found to play a role in regulating endothelial exocytosis. Specifically knockdown of *STXBP5* with an siRNA increased VWF and p-selectin exocytosis, a finding that was supported by *in vivo* studies, where mice lacking *Stxbp5* had higher levels of plasma VWF and increased platelet-endothelial interactions (Zhu et al., 2014). Conversely, a study published at the same time, implicated *STXBP5* as a positive regulator of platelet granule exocytosis (Ye et al., 2014). Therefore *STXBP5* was shown to have distinctly different roles in these two cell types and highlights the importance of understanding cell specific functions of proteins.

Other loci that were not identified through GWAS studies have been identified via knowledge of their function and hypothesising a role in VWF levels. Via this approach, gain of function single nucleotide polymorphisms (SNPs) have been identified in the arginine vasopressin 2 receptor (*V2R*) gene (*AVPR2*). In vascular endothelial cells, stimulation of *V2R* causes WPBs to release their contents, thus increasing levels of VWF and FVIII in the plasma. Gain of function SNPs such as

the 12E mutation in V2R (12E-V2R) demonstrate increased binding of its ligand AVP and thus increased signal transduction leading to increased VWF and VIII release into the circulation (Nossent et al., 2010).

This approach has also shed light on aspects of the VWF clearance pathway from the circulation. VWF clearance is still not fully understood, although macrophages within the liver appear to play a key role in this process (van Schooten et al., 2008). *In vivo* experiments using mice that were deficient in the macrophage lipoprotein receptor LRP1, showed increased basal levels of VWF and also significantly increased VWF half-life, thus demonstrating a role for this receptor in VWF uptake (Rastegarlar et al., 2012). Studies on two type 2B mutants (p.Arg1306Gln and p.Val1316Met) have since shown that the A1 domain of VWF contains an LRP1 binding site. LRP1 binding is enhanced in these two mutations leading to increased clearance (Wohner et al., 2015).

The fact that seemingly very significant loci have not been identified via GWAS studies raises the possibility that other genetic loci that directly influence VWF levels are still unidentified. Indeed, using linkage analysis based on sibling structure within two large cohorts, Desch et al identified a locus at 2q12-2p13 (Desch et al., 2013). This locus was found to be similarly important in its contribution to the variance in VWF levels as the *ABO* locus (19% and 25% respectively).

In 2016, a GWAS study was undertaken to identify genetic loci specifically associated with low VWF levels (van Loon et al., 2016). This analysis identified familiar loci from previous GWAS studies (namely *ABO*, *VWF*, *STXBP5* and *STAB2*), however a novel locus of the ubiquitin fold modifier 1 (*UFM1*) gene was also identified. UFM1 is a protein conjugating system thought to function in cellular homeostasis and ER function. In pancreatic beta cells, ER stress has been shown to induce expression of UFM1 (Lemaire et al., 2011) and it has been suggested that UFM1 is a target for the ER stress induced transcription factor XBP1 (Zhang et al., 2012).

Although not directly associated with VWF levels, one study has reported a potential role for the autophagy pathway in VWF secretion (Torisu et al., 2013). Inhibition or knockdown of key autophagy genes inhibited VWF secretion *in vitro* and an endothelial specific *Atg7* knockout mouse model showed decreased VWF

release and a reduction in HMW multimers following epinephrine stimulation. Inhibition of autophagy also resulted in smaller and rounder WPB.

Despite the association of a number of other loci with VWF and FVIII levels, there are still relatively few studies that demonstrate the molecular pathogenicity of these candidates. In this respect some of these candidate loci are vague associations and further work is required to uncover their involvement, if any, in VWF level modulation.

1.9 Ethnicity effect

When attempting to assign pathogenicity to any given variant identified in *VWF*, it is important to consider what constitutes normal genetic variation within the gene particularly across different ethnic groups. Some mutations identified as candidate causative mutations in European patients have now been shown to occur with allele frequencies of up to 18% in African Americans (Bellissimo et al., 2012).

Studies such as the 1000 Genomes Project have utilised next-generation sequencing technologies to document sequence variants in a diverse group of ethnic backgrounds (The Genomes Project, 2012, The Genomes Project, 2015). This should help to establish a database of ethnic specific sequence variations and prevent the incorrect classification of neutral variants as pathogenic mutations (Wang et al., 2013b). More recently the Exome Aggregation Consortium (ExAC) has published exome DNA sequence data for over 60,000 individuals from diverse ancestries (Lek et al., 2016), providing a valuable reference data set for the medical interpretation of genetic variants, accessible via the ExAC database (<http://exac.broadinstitute.org/>). While it is clear that the *VWF* gene is ethnically diverse, further work is required to determine the contribution of ethnicity to disease.

1.10 Hypotheses and research aims

The Sanger sequencing approach of previous studies are unreliable for the detection of large deletions or duplications in heterozygotes where the normal allele will mask the absence of the deleted allele. Recent findings from MLPA

based studies have shown that CNV contributes to the mutation spectrum of type 1 VWD, where large in-frame exonic deletions disrupt the VWF protein leading to dysfunction and disease (Sutherland et al., 2009, Hampshire et al., 2010, Cartwright et al., 2013). However, a major limitation of these studies and others is that they are limited to coding regions of the *VWF* locus. Therefore non-coding intronic and 5' and 3' intergenic regions remain unexplored in terms of CNV analysis. Furthermore, the nature of MLPA probe design leaves open the possibility that smaller CNV that may overlap intron-exon boundaries will not be detected due to lack of coverage. Array comparative genomic hybridisation (aCGH) is an alternative technique that could provide a more comprehensive CNV screen of the *VWF* locus. The applicability of aCGH to CNV analysis in *VWF* has been demonstrated previously in an analysis of type 3 VWD patients (Bellissimo et al., 2011). However the array design used in this study omitted probe coverage in the *VWF* pseudogene region and intergenic flanking regions.

Previous studies that identified exonic in-frame deletion CNV in type 1 VWD patients show reduced secretion *in vitro*. However, the pathogenic mechanism of this reduced secretion and the reduced VWF levels observed in patients remains to be elucidated.

1.10.1 Hypotheses

In this study the following experimental hypotheses were investigated:

- i. A subset of CNVs within *VWF* are undetectable via MLPA and sequencing based analyses and these CNV contribute to the VWD mutation spectrum.
- ii. Non-coding intronic and intergenic CNV in *VWF* contribute to VWD.
- iii. Exonic CNV consisting of large in-frame heterozygous deletions cause reduced VWF secretion due to ER retention of the mutant protein leading to impaired WPB biogenesis and secretion.

1.10.2 Aims

The overall aims of this study therefore were two-fold:

Firstly to design a custom *VWF* microarray that could be used to perform a detailed CNV screen of the entire *VWF* locus including intronic and intergenic flanking regions.

Secondly, to carry out a thorough *in vitro* functional characterisation of three previously identified exonic deletion CNV from a type 1 VWD cohort.

These studies aimed to contribute to the current understanding of CNV prevalence in *VWF* and its pathogenicity in VWD.

2 Materials and Methods

2.1 Materials

General laboratory chemicals were obtained from Sigma-Aldrich Company Ltd. (Dorset, UK) or BDH laboratory supplies (Lutterworth, Leicestershire, UK) unless otherwise stated. Solutions were prepared using deionized water and where necessary, were sterilised by autoclaving.

2.1.1 List of stock solutions

Ampicillin (100 mg/ml)

Bovine serum albumin (BSA) (1% w/v made fresh each use)

Dimethyl sulphoxide (DMSO; 100%)

Ethanol (100%)

Ethidium Bromide (10 mg/ml), used at 1mM final concentration

Ethylenediaminetetraacetic acid (EDTA; 0.5 M)

Foetal bovine serum (100%) was obtained from Gibco, Life Technologies, Paisley, UK.

Glycerol (100% (v/v))

Goat Serum (100%), used at 5% (v/v)

Isopropanol (100%)

Phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 0.7 mM KH₂PO₄,
4

mM Na₂HPO₄)

Sodium chloride (NaCl; 5 M)

Sodium acetate solution (3M CH₃COONa or abbreviated to NaOAc)

Tris/Borate/EDTA (TBE) buffer (5x: 0.45M Tris, 0.45M Boric Acid and 0.01M EDTA (0.5x TBE working solution: 0.045M Tris, 0.045M Boric Acid and 0.001M EDTA)

Tris/EDTA (TE) buffer (10 mM Tris-HCl, 1 mM disodium EDTA, pH 8.0)

Trypsin-EDTA (1x)

Triton X-100 (100%)

2.1.2 Microbiology reagents

The NM554 *E.coli* strain was originally obtained from Agilent Technologies (Cheshire, UK)

2.1.2.1 Plasmids and VWF expression vectors

VWF cDNA previously cloned into the pcDNA™3.1 (-) plasmid vector (VWF-pcDNA3.1; (Life Technologies Ltd., Paisley, UK)) was a gift from Prof. Reinhard Schneppenheim (Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany). The pCIneo-VWF expression vector was kindly provided by Dr. McKenzie Bowman (Queens University, Canada).

The pGEM-T vector used for TA cloning of the VWF 5' deletion PCR product during breakpoint mapping was obtained from Promega (Southampton, UK).

2.1.2.2 Bacterial culture media

Luria Bertani (LB) agar (37 g/l) (Merck, Dormstadt, Germany)

LB-ampicillin agar (LB agar supplemented with 100 µg/ml ampicillin)

LB broth (25 g/l) (Fisher Bioreagents, Loughborough, UK)

LB-ampicillin broth (LB broth supplemented with 100 µg/ml ampicillin)

NZ amine (Casein hydrosylate) yeast extract (NZY) broth were obtained from Fisher Scientific UK Ltd. (Loughborough, UK).

Super Optimal broth with Catabolite repression (SOC) media

2.1.2.3 Plasmid purification reagents and kits

Plasmids were extracted from transformed, overnight bacterial cultures using the QIAprep Spin Miniprep kit (small scale purification) or the EndoFree Plasmid Maxi kit (QIAGEN, Crawley, UK). All reagents were provided with the kit.

2.1.2.4 TA cloning reagents

The pGEM-T Easy Vector System obtained from Promega (Southampton, UK) was used for TA cloning. Reagents for the addition of A overhangs included: 5x GoTaq PCR buffer, 2mM dATP and GoTaq polymerase. All were obtained from Promega.

2.1.2.5 Mutagenesis reagents

The QuikChange Lightning Site Directed Mutagenesis kit was obtained from Agilent Technologies (Cheshire, UK). This kit included XL10-Gold Ultracompetent cells.

2.1.3 General molecular biology reagents

2.1.3.1 PCR and agarose gel electrophoresis

Agarose and DNA ladders were obtained from Biorun (Biorun reagents limited, London, UK). GoTaq DNA polymerase was obtained from Promega (Southampton, UK). OneTaq Hot Start 2x Mastermix with standard buffer was obtained from New England Bioscience (NEB) (Hitchin, UK). Reddymix was obtained from Fisher Scientific. (Loughborough, UK). Expand-Long template system was purchased from Roche (Burgess Hill, West Sussex, UK). The QIAquick gel extraction kit was obtained from QIAGEN (Crawley, UK). The Human DNAOK! DNA integrity assay was obtained from Microzone (Haywards Heath, UK).

2.1.3.2 PCR, genotyping, mutagenesis and sequencing primers

Primer sequences were designed using Primer 3 as discussed in section 2.2. Primers were obtained from Eurofins MWG Operon (Ebersberg, Germany). Lyophilised primers were resuspended in molecular grade deionised water (dH₂O), according to the supplier's recommendations to generate a stock concentration of 100 pmol/μl (100 μM). This was diluted 1:10 to generate a 10μM working stock.

Table 2.1 shows the 5' intergenic region deletion primers used to map the 5' deletion breakpoints and used for genotyping assays.

To generate the exon 32-34 deletion in the VWF-pCI-neo vector, the primers shown in table 2.2 were used. Following mutagenesis, internal sequencing primers were used to sequence across the VWF cDNA insert (Table 2.3).

2.1.4 Array Comparative Genomic Hybridisation (aCGH) reagents

Microarrays and associated labelling, hybridisation and wash buffer kits and components were obtained from Agilent Technologies (Cheshire, UK).

Table 2.1 VWF 5' intergenic region deletion primers

Primer	Sequence 5' → 3'
5'del-I_F (FAM)	GGCAATAGAGAGTGATATGGAAAGA
5'del-I_R	GTGCCAGAGTGTTCATATGAGTTACA
5'del-II_F (VIC)	AACAAATTAAGTATACACGCAACGT
5'del-II_R	AAATACCTTCATTGCTTCTTGACCT
5'del-I_F_Int1 (FAM)	TTTTTGTTTGTCTTATCTCCTCCAG
5'del-II_F_Int2 (VIC)	AGGTGCAAACCTTTCTGATAATTGAA

Table 2.2 Mutagenesis primers used to generate rVWF32-34del

Primer	Sequence (5' – 3')
rVWF32-34del F	GATGCCGCCAGGTCCAACAGCGTGTGCACAGGCAGCTCC
rVWF32-34del R	GGAGCTGCCTGTGCACACGCTGTTGGACCTGGCGGCATC

Table 2.3 Internal sequencing primers for VWF cDNA sequencing

VWF internal sequencing forward primer	Sequence (5' – 3')
T7F	TAATACGACTCACTATAGGG
1F	CGGCAACTTTCAAGTCCTGCTGT
2F	CAGTGTGCCCTGCTGGTATG
3F	CAGCCTTGTGAAACTGAAGCATGG
4F	CGAAAGGCCAGGTGTACCTGCAG
5F	CTGGAGTGCATGAGCATGGGCTG
6F	CATCATTCTGCTGCTGGGCAAAG
7F	CAGAGCTGCGAGGAGAGGAATC
8F	CTTTGTGGTGGACATGATGGAGC
9F	CATGGCACAAGTCACTGTGGGC
10F	CTAATGCCAACGTGCAGGAGCTG
11F	TCATCCTGGTCACGGACGTCTC
12F	CAAGCTGACTGGCAGCTGTTCTT
13F	CATCCTGGAGGAGCAGTGTCTTG
14F	CACGTGTGGCCTGTGTGAAGTAG
15F	CAGAAGCCCTGTGAGGACAGCT
16F	CAAGGAAGAAAATAACACAGGTG

2.1.5 Mammalian cell culture and transfection

2.1.5.1 Human Umbilical Vein Endothelial Cell (HUVEC) culture

Human umbilical vein cells (HUVECs) were obtained from healthy donors according to the University of Sheffield research ethics guidelines. HUVECs were cultured in endothelial cell culture medium (ECM) obtained from Sciencell (Caltag Medsystems, Buckingham, UK).

2.1.5.2 Human Embryonic Kidney (HEK) 293 Cell culture

HEK293 cells were a kind gift from Dr. Kurt de Vos (University of Sheffield) and were originally obtained from the American Type Culture Collection (ATCC, Rockville, USA).

Dulbecco's Modified Eagles Media (DMEM) cell culture media, foetal bovine serum (FBS) were obtained from GIBCO, Life Technologies (Paiseley, UK).

2.1.5.3 HEK 293 cell freezing media reagents

FBS obtained from GIBCO (Life Technologies, Paisley, UK) and DMSO (Sigma).

2.1.5.4 Transfection reagents

Lipofectamine LTX plus was obtained from Invitrogen, (Life Technologies, Paiseley, UK). The X-fect reagent was obtained from Takara Clontech (Saint-Germain-en-Laye, France).

2.1.5.5 Enzyme Linked Immunosorbent Assay (ELISA) and related reagents

5x passive lysis buffer was obtained from Promega (Southampton, UK). Protease inhibitor cocktail tablets were supplied by Sigma.

Capture and detection anti-VWF antibodies were supplied by Enzyme Research Laboratories (Swansea, UK).

Enzyme Linked Immunosorbent Assay (ELISA) buffers, reagents and ELISA standard VWF:Ag calibrator were obtained from Affinity Biologicals (Quadrantech Diagnostic Ltd, Epsom, Surrey, UK). ELISA buffers included in the kit contained the following: Coating buffer (50 mM carbonate), PBS, wash buffer (PBS-Tween (0.1%,v/v)), blocking buffer (PBS-BSA (1%, w/v)), sample diluent (100mM HEPES, 100mM NaCl, 1% w/v BSA, 0.1% v/v Tween-20, pH7.2), substrate buffer (27mM Citric acid, 97mM Na₂HPO₄, pH5.0) and OPD substrate: (o-Phenylenediamine.2HCl). 6% (w/v) hydrogen peroxide (H₂O₂) was also required for the OPD reaction.

2.1.6 Immunocytochemistry and fluorescent microscopy reagents

BD Cytotfix solution, containing 4.2% formaldehyde (w//w) was obtained from BD Bioscience (Oxford, UK).

Quenching solution consisting of 50mM NH₄Cl was used to quench excess formaldehyde. Blocking solution consisted of 1% BSA and 5% normal Goat serum in PBS. All reagents were obtained from Sigma.

2.1.6.1 Antibodies

Primary antibodies for immunofluorescent microscopy were purchased as follows: polyclonal rabbit anti-human VWF (Dako UK Ltd, Cambridgeshire, UK), mouse anti-human calnexin (Abcam, Cambridge, UK), sheep anti-human TGN46 (AbD Serotec, Kidlington, UK) and mouse anti-Rab27a (Santa Cruz, USA). All secondary antibodies were purchased from Life Technologies (Paiseley, UK) and included goat anti-rabbit IgG Alexa Fluor 488, goat anti-mouse IgG Alexa Fluor 555, goat anti-mouse IgG Alexa Fluor 568 and donkey anti-sheep IgG Alexa Fluor 647.

2.1.6.2 Non-immunological stains

The actin cytoskeleton stain Phalloidin was obtained from Sigma.

DAPI (4',6-diamidino-2-phenylindole) was obtained from Life Technologies and was prepared as a 1000x stock solution and stored at -20°C.

2.1.6.3 Mounting media

Prolong Gold mounting medium was obtained from Life Technologies (Paisley, UK). Vectashield anti-fade mounting medium was obtained from Vector Laboratories Ltd., (Peterborough, UK).

2.1.7 Patient cohort DNA samples

DNA samples were originally obtained as part of the large European study, Molecular and Clinical Markers for the Diagnosis and Management type 1 von Willebrand Disease (MCMDM-1VWD), aimed at determining the relationship between patient phenotype and genotype in type 1 VWD (Goodeve et al., 2007). Samples were recruited from index cases (IC), affected family members (AFM), unaffected family members (UFM) and healthy controls (HC), from 12 centres in 9 EU countries. Genotypic and phenotypic data was provided with these samples including VWF antigen (VWF:Ag), VWF ristocetin cofactor activity (VWF:RCo) and FVIII coagulant activity (FVIII:C). All samples were obtained following ethics committee review, where locally required and following informed consent for study participation. All experimental methodologies were undertaken in accordance with the Helsinki Declaration.

In addition to the EU cohort, DNA samples were provided from Canada and the US Zimmerman Program for the Molecular and Clinical Biology of VWD. Initially, samples were requested from patients with previously confirmed CNV. Some of these samples were used as part of the validation for the microarray studies (see chapter 3 section 3.3.3). 18 individual samples were provided for validation purposes, consisting of EU, Canadian, UK and patients recruited to the Zimmerman Program based in Milwaukee.

Additionally, 10 samples were obtained from type 3 patients, recruited to the Type 3 von Willebrand International Registries Inhibitor Prospective Study (3WINTERS-IPS). In 2013, the 3WINTERS study commenced with the aim of

determining the genetic basis of type 3 VWD. A subset of these samples were sent to Sheffield for aCGH analysis.

In the present study, a total of 84 test patients were provided for aCGH analysis. These consisted of 55 type 1 VWD individuals (from 47 families), 6 type 2 VWD patients and 23 type 3 patients. Test patients were selected on the basis that they had either no previously reported mutation or CNV that was considered causative of VWD.

All control DNA was purchased from Agilent (Agilent Technologies, UK). These male and female DNAs were extracted from single genotyped individuals.

2.1.8 Laboratory equipment

Accuspin-micro Microfuge (Fisher Scientific, UK).

Beckman J2-21M/E Ultracentrifuge

Nanodrop ND-1000 spectrophotometer (Labtech International, East Sussex, UK).

PCR thermal cyclers:

Veriti 96 well thermocycler (Applied Biosystems).

GeneAmp PCR System 9700 (Applied Biosystems).

Bio-Rad Gel Doc XR Gel Documentation System (Hemel Hempstead, UK).

Agilent hybridisation oven and Agilent microarray scanner (model G2505C) from Agilent Technologies (Cheshire, UK)

Varioskan Flash Reader (Thermo Scientific, Kent, UK).

Nikon TiE inverted microscope with a 100x oil objective and Andor Zyla sCMOS camera was from Nikon UK LTD (Kingston Upon Thames, Surrey, UK).

Deltavision OMX SIM microscope from GE Healthcare Life Sciences (Buckinghamshire, UK)

2.1.9 Image Analysis

Image analysis packages included ImageJ/FIJI and IMARIS software for 3D reconstructions.

2.2 Methods

2.2.1 Array Comparative Genomic Hybridisation (aCGH) introduction

The development of array based comparative genomic hybridisation (aCGH) has enabled the analysis of copy number variation (CNV) across the entire genome and its role in numerous diseases to be explored (Pinkel and Albertson, 2005). Indeed it is now considered the gold standard by diagnostic laboratories for the detection of CNV and is widely used for the genetic diagnosis of developmental delay disorders and congenital abnormalities (Miller et al., 2010). It is a comparative approach as it involves co-hybridisation of differentially labelled DNA samples to a microarray containing numerous oligonucleotide probes specific to genomic regions of interest. A summary of the aCGH protocol is provided in figure 2.1 briefly, patient DNAs are labelled with Cyanine 3 (Cy3)-deoxyuridine triphosphate and control DNA is labelled with Cyanine 5 (Cy5)-dUTP. Differentially labelled DNAs are then combined and hybridised to the microarray. Following hybridisation, the arrays are washed and scanned. Scan data is then processed using feature extraction software, which generates Cy3/Cy5 ratios for all probe positions. This data can then be interpreted in CNV analysis software.

2.2.1.1 DNA quantification and quality analysis for use in aCGH

DNA samples were quantified using a NanoDrop spectrophotometer ND-1000 (Fisher Scientific). DNA quality was assessed by recording the 260/280 ratio, as a measure of contaminating proteins and the 260/230 ratio, as a measure of any organic solvent contamination such as phenol used in some DNA extraction protocols. A 260/280 of >1.8 and a 260/230 of >2 indicates DNA free from any contaminants that may interfere with or inhibit downstream processes. It was necessary to have ~3 µg of DNA for each sample to enable the clean-up procedure described in 2.2.2.

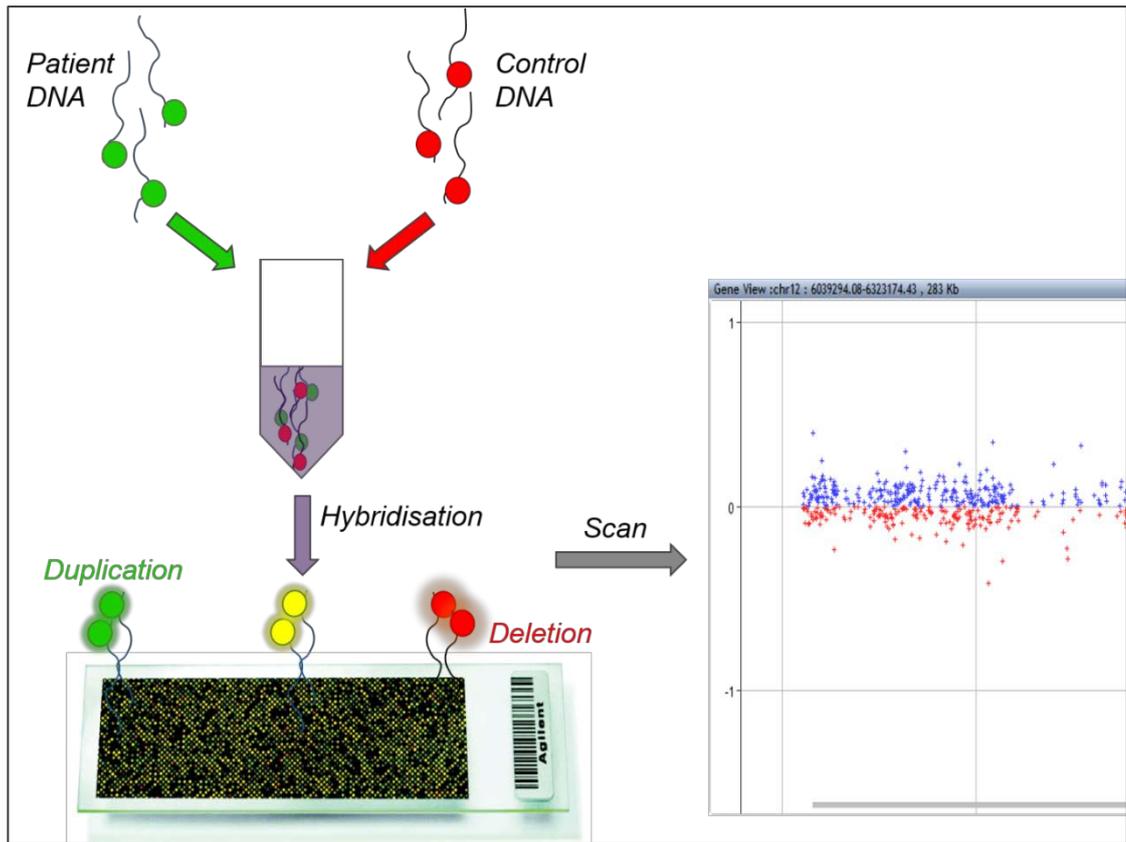


Figure 2.1. Overview of aCGH. Patient and control DNAs are differentially labelled with Cy3 and Cy5 dyes. These are then combined and hybridised to the microarray, consisting of numerous oligonucleotide probes. Following hybridisation, the microarray is washed, and scanned and the log₂ ratio of Cy3 to Cy5 is calculated via feature extraction and analysis software

2.2.1.2 Sodium acetate and ethanol precipitation DNA clean-up

All DNA samples that were analysed via aCGH were cleaned up via this method. 3µg of DNA for each sample was resuspended in TE buffer to a final volume of 50µl. One tenth of this volume, i.e. 5µl of 3M sodium acetate was then added to the sample and gently mixed by flicking the tube. Next, 2.5x of this 55µl volume i.e. 137.5µl of 100% ethanol was added to precipitate the DNA. The tube was inverted multiple times to mix and enable the DNA to precipitate. Samples were then left at -20°C overnight to allow complete precipitation. Samples were then centrifuged in a bench top microfuge at 13,000 g for 15 min in order to pellet the DNA. Supernatants were carefully removed by pipetting. DNA pellets were then washed with 300 µl 70% ice cold ethanol and centrifuged for a further 13,000 g for 5 min. The supernatant was removed as above and after a pulse spin to remove excess ethanol, the DNA pellet was left to air dry for ~5-10 min before resuspending in ~20 µl 1x TE buffer. DNA quantity was then reanalysed along with A260/280 and A260/230 ratios.

2.2.1.3 DNA integrity analysis

High quality intact high molecular weight DNA was required for aCGH analysis. All samples for aCGH analysis were therefore subjected to a DNA quality check using the Human DNAOK! Assay kit (Microzone Limited, Haywards Heath, UK). Briefly, a multiplex PCR reaction was performed, amplifying 6 different regions of the human genome. Good quality DNA should produce 6 amplicons whereas degraded or poor quality DNA may result in the loss in one or more of these amplicons, reflecting the loss of intact, high molecular weight genomic DNA. These assays were kindly performed by Dr. Peter Winship (Sheffield Diagnostic Genetics Service).

2.2.1.4 Custom VWF microarray design

The custom microarray was based on the 8x15k platform which uses 15,000 60-mer oligonucleotide probes, spotted onto each of 8 sub-arrays. Arrays were designed using SureDesign software v1.3 (Agilent) using genome build hg19, GRCh37. Briefly, a standard probe set was generated from the VWF genomic coordinates (GRch37/hg19). Further probe groups were then generated using coordinates from the 5' and 3' flanking sequences, up to the genes flanking *VWF*, *CD9* and *ANO2* respectively. A further VWF exonic probe group was kindly provided by Dan Bellissimo (Blood Research Institute, BloodCenter of Wisconsin, Milwaukee, WI, USA) and Madhuri Hegde (Emory Genetics Laboratory, Department of Human Genetics, Emory University, Atlanta, GA, USA) (Bellissimo et al., 2011). The pseudogene region (exons 23 – 34) was designed with altered probe selection parameters to allow increased probe coverage. For these coordinates, the similarity filter was removed to allow for the presence of the partial pseudogene on chromosome 22. Any exonic gaps that remained were filled using the genomic tiling option in SureDesign.

Finally, all the probe groups were saved as a bed file and imported into the UCSC browser (*H. sapiens*, hg19, GRCh37, February 2009) as a custom track. A manual final check was then undertaken by scanning across the region, checking for uneven probe coverage or areas of significant probe overlap. Any probes that appeared to be completely overlapping others were deleted from the bed file before finalising the array design. All probes were represented in triplicate and further Agilent probe groups were added from across the genome for normalisation purposes

2.2.1.5 Array design iterations and modifications

Further array design iteration were carried out in an attempt to improve probe coverage in and around the *VWF* pseudogene region, particularly exon 34 and intron 34. These re-designs are discussed in detail in chapter 3.

2.2.1.6 Fluorescent labelling of DNA for aCGH

DNA samples selected for aCGH analysis were labelled using an enzymatic, random primer based approach as described in the Agilent aCGH manual. Briefly, 500ng of cleaned patient or control DNA in a final volume of 11µl was added to a PCR tube. 2.5µl of random primers was then added and the samples were denatured in a PCR machine at 95°C for 10 min in order to denature the DNA and enable partial fragmentation. Samples were then rapidly cooled by immediately placing on ice for 5 min. This snap-cooling method prevents re-annealing of the template DNA and results in increased labelling efficiency. A labelling master mix was prepared consisting of 42 µl 5x reaction buffer, 21 µl dNTPs, 12.6 µl Cy3-dUTP or Cy5-dUTP and 1.0 µl exo-Klenow fragment, before mixing well. Exo-Klenow is a large fragment produced from proteolysis of DNA polymerase I that lacks 5' to 3' exonuclease activity (Klenow and Henningsen, 1970). Next, 9.5 µl of the Cy3 or Cy5 labelling master mixes was then added to each patient sample DNA or control DNA respectively and mixed well by pipetting. Samples were then transferred to a thermal cycler machine and incubated at 37°C for 2 h followed by a 10 minute deactivation step at 65°C.

2.2.1.7 Clean-up of fluorescently labelled DNAs

Labelled samples were transferred to a microcon filter (Agilent, UK) containing 430 µl TE buffer, in a 2 ml collection tube. Samples were centrifuged at 13,000 xg in a bench top microfuge for 10 min. The flow through was discarded and 480 µl TE buffer was added to the column before centrifuging for a further 10 min at 13,000 g. The flow through was discarded again and the column was then inverted into a clean 2 ml collection tube and centrifuged at 1000 g for 1 min to elute the labelled DNA. Samples were then vacuum concentrated at high heat for ~20 min. Dried labelled DNAs were then resuspended in 9.5 µl 1x TE buffer and left to reconstitute at room temperature for 1 h in the dark. Labelling figures were measured using a NanoDrop in the microarray mode. DNA concentrations and Cy3 and Cy5 figures were recorded for patient and control DNAs respectively.

2.2.1.8 Hybridisation

Labelled DNAs were combined with labelled control DNA. A hybridisation master mix consisting of COT1 DNA, Agilent hyb blocking solution and hybridisation buffer was generated and added to the labelled combined DNA. Samples were then incubated at 95°C in a hot block for 3 min before placing in a water bath at 37°C for 30 min. Samples were briefly centrifuged to collect all the contents of the tube before pipetting 40 µl on to the array gasket slide. After all eight gaskets were loaded the array was overlaid and the gasket was clamped tightly. Arrays were hybridised at 65°C for 24h at max RPM rotation in the Agilent hyb oven.

2.2.1.9 Array washing, scanning and feature extraction

The arrays were then scanned at 3µm resolution in an Agilent scanner. TIF images were feature extracted using the Agilent FE software v11.5 which normalises cy3 and cy5 fluorescence for each probe before calculating their ratio and provides a number of quality control (QC) parameters (Table 3.6). Calculated QC parameters include signal to noise ratios (cy3 and cy5) and signal intensities for each. The cy3:cy5 ratio is expressed as a log₂ ratio.

2.2.1.10 Feature extraction, data analysis and aberration detection

All aCGH data was analysed using Agilent Cytogenomics software (version 2.7.22.0) with the ADM1 aberration detection algorithm. Initial analyses used the default CGH V2 detection algorithm, however this excluded probes in the pseudogene region. Removal of the design level filter and any similarity filters allowed more probes in the pseudogene region to be used in the analysis. This new analysis protocol was saved as Trial-method-1.

Feature extracted txt-files were then imported into Agilent CytoGenomics software v2.7 for copy number analysis. Imported aCGH data was analysed initially using the Default Analysis Method – CGH v2, which uses the ADM2 algorithm. GC correction normalisation and diploid peak centralisation were turned on. Reference DNA was re-labelled to red (cy5). In order to improve CNV detection and calling in the pseudogene region, subsequent analyses were performed using a modified analysis method in which the design level filter was

removed. This removed the homology filter from the analysis and allowed probes with homology to other regions to be included in the analysis.

2.2.2 Polymerase Chain Reaction (PCR)

PCR is a commonly used technique in molecular biology that enables the oligonucleotide primer based amplification of a short stretch of DNA via cyclic reaction consisting of denaturation of DNA, primer annealing and extension followed by a final extension.

2.2.2.1 Primer design

All primers used in this study were designed using the web based primer design tool Primer 3 (<http://primer3plus.com/cgi-bin/dev/primer3plus.cgi>) (Untergasser et al., 2012).

2.2.2.2 Standard PCR

Standard PCR amplification of short amplicons was performed using ReddyMix following the manufacturer's guidelines. Briefly, 22.5 μ l ReddyMix (containing 0.625 u of ThermoPrime *Taq*, 75 mM Tris-HCl (pH 8.8), 20 mM $(\text{NH}_4)_2\text{SO}_4$, 0.01% Tween 20, 1.5 mM MgCl_2 and 0.2 mM dNTPs), 1.0 μ l dH_2O , 0.5 μ l forward and reverse primers (at 10 pmol/ μ l) and 0.5 μ l DNA (100 ng/ μ l) were added to a PCR tube per 25 μ l reaction.

Thermal cycling conditions for standard PCR were based on the following: denaturation at 94-96°C for 2 min, then 35 cycles of 95°C for 25 s, annealing 50°C-65°C for 30 s to 1 min, extension at 72°C for 1 min and a final extension of 72°C for 5 min.

2.2.2.3 OneTaq Hot-Start PCR

For longer amplicons (up to ~5 Kb) and regions that proved problematic for standard PCR approaches, the OneTaq hot-start system was used. This system can reduce non-specific amplification as the polymerase is inactivated at lower

temperatures. In this study, the OneTaq hot start 2x master mix was used following the manufacturer's recommendations. Briefly, a 25 μ l reaction consisted of: 12.5 μ l OneTaq Hot Start Master Mix (containing 20 mM Tris-HCl (pH8.8), 22 mM KCl, 0.2 mM dNTPs, 1.8 mM MgCl₂, 5% glycerol, 0.05% Tween-20, 25 units/ml OneTaq DNA Polymerase, 22 mM NH₄Cl, and 0.06% IGEPAL CA-630), 10.5 μ l dH₂O, 0.5 μ l each of forward and reverse primers (at 10 pmol/ μ l), and 1.0 μ l DNA (100 ng/ μ l). The following thermal cycling parameters were used: initial denaturation at 94°C for 30 sec followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 58°C for 60 sec and elongation at 68°C for 6 min, then a final extension at 68°C for 5 min.

2.2.2.4 Breakpoint mapping

Using the approximate maximum aCGH defined breakpoints, the Ensembl database was used to export the DNA sequences of desired regions. Regions were exported with an additional 1000bp of flanking sequence both 3' and 5' of the deletion breakpoints. Sequences were then pasted into Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>), and used as the source sequence for primer design. Primers were designed to have an optimum size of 25 nt and a minimum T_m of 55°C. Products ranging from 5kb to 7kb were selected in order to generate amplicons that contained the WT sequence. Primer sequences were then checked via BLAST, to check specificity for the region of interest. A long range PCR strategy was employed to enable amplification of both the large WT non-deleted alleles and the smaller deletion specific product, using the Expand Long Template PCR system (Roche).

2.2.2.5 Expand long template PCR protocol

This kit contains both a thermostable *Taq* polymerase and a thermostable DNA polymerase with proof reading activity, thus preventing *Taq* polymerase fall-off due to base mis-incorporations. In summary, long range PCR reaction mixes were made up in a total volume of 50 μ l consisting of 1.5 μ l each of forward and reverse primers (10 μ M stock, giving 300 nM final), 1.75 μ l dNTPs (10 mM stock, 350 μ M in reaction), 5 μ l Expand Long template buffer 1, 1 μ l DNA (~100-500

ng/μl), 0.75 μl Expand long template enzyme mix (3.75 U) and 36.75 μl ddH₂O. Long range PCR reactions were performed using the manufacturer's recommended thermal cycling parameters outlined in table 2.4 This protocol included a gradual increase in the extension time in order to ensure a higher yield of amplification products.

Table 2.4 Thermal cycling parameters for Expand Long Template PCR

Reaction step	Temperature	Duration	Cycles
Initial denaturation	94°C	2 min	1
Denaturation	94°C	10 s	10
Annealing	55-65°C (59)	30 s	
Elongation	68°C	6 min	
Denaturation	94°C	15 s	15-25
Annealing	55-65°C (59)	30 s	
Elongation	68°C	6 min + 20 s cycle elongation for each successive cycle	
Final elongation	68°C	7 min	1
Cooling	4°C	-	1

2.2.2.6 Multiplex PCR

Multiplex PCR is a variation on standard PCR that allows the amplification of multiple targets or amplicons. In this study multiplex PCR was performed using the QIAGEN Multiplex PCR kit. This kit contains a multiplex master mix with pre-optimised concentrations of HotStart Taq DNA Polymerase and MgCl₂, dNTPs,

and PCR buffer that contains the novel factor MP. The molar concentrations of components in this proprietary mix or the identity of novel factor MP are not provided by the supplier. Multiplex PCR reactions were set up as 25 µl reactions consisting of: 12.5 µl QIAGEN multiplex master mix, 1 µl of each reverse primer, 0.5 µl of the multiple forward primers and 7.5 µl nuclease free H₂O. The thermal cycling parameters were as follows: initial denaturation at 95°C for 15 min, followed by 30 cycles of 94°C for 30 s, 52°C-66°C for 90 s annealing, 72°C extension for 1 min and a final extension at 72°C for 10 min.

2.2.2.7 Multiplex genotyping

The forward primers shown in table 2.1 were tagged with 6-carboxyfluorescein (6-FAM) and/or VIC labels to allow high throughput genotyping using the ABI 3730 DNA analyser. The LIZ1200 size standard was used to determine amplicon sizes. This analysis enabled high throughput of multiplex analyses as amplicons were detected via fluorescence and samples could be set up in 96-well PCR plates.

2.2.2.8 DNA sequencing and sequence analysis

All sequencing was performed by the Core Genomic Facility at the University of Sheffield, using the ABI 3730 automated DNA analyser. This uses the Sanger sequencing method of fluorescent, chain terminating dideoxynucleotide incorporation. DNA sequence chromatograms were analysed using Finch TV Software, Version 1.4 (Geospiza, Seattle, USA), for initial sequence analysis.

2.2.3 General molecular biology methods

2.2.3.1 Agarose gel electrophoresis of DNA

Agarose was dissolved in 0.5x TBE buffer by boiling. Gels were made up at different concentrations depending on the size of DNA products that were being electrophoresed. Therefore agarose gel concentrations ranged from ~0.7% (w/v) for larger products and up to 1.5% (w/v) for smaller products. Ethidium bromide

was added at a final concentration of 1mM after the molten agarose had cooled to ~60°C. Gels were cast in horizontal gel trays and once set, were submerged in 0.5x TBE in a suitable size electrophoresis tank and electrophoresed at 80-120 V for 30 min to 1 h depending on the size of DNA products that were to be resolved. Gels were visualised using a Bio-Rad Gel Doc XR Gel Documentation System.

2.2.3.2 QIAquick gel extraction

The PCR product was purified using the Qiagen QIAquick PCR purification kit according to the manufacturers recommended protocol. In brief, the PCR product band of interest was excised from the agarose gel and placed in a 1.5 ml tube. The gel slice was weighed and then 3 volumes of buffer QG was added. The gel slice was melted via incubating at 50°C in a hot block with intermittent vortexing and mixing for 10 min. One gel volume of isopropanol was added and the sample was vortexed. Next, up to 800 µl of the sample (depending on sample volume) was added to a QIAquick spin column and centrifuged at 17900 g for 1 min. Flow through was discarded and the centrifugation step was repeated for samples larger than 800 µl. After all the sample volume had passed through the spin column, 500 µl of buffer QG was added and the above centrifugation was repeated. The spin column was then washed by adding 750 µl of buffer PE and centrifuging as above. The flow through was discarded and the centrifugation was repeated to ensure removal of all residual ethanol. The end of the column was dabbed onto absorbent filter paper to ensure no flow through remained before the column was placed into a clean 1.5 ml centrifuge tube and DNA was eluted via the addition of 20 µl of buffer EB. To ensure maximum recovery of DNA, the EB buffer was left on the spin column filter for 5 min prior to centrifugation at 17900 g to elute the purified DNA. Purified DNA was stored at 4°C prior to quantification or downstream application.

2.2.3.3 TA cloning

Desired PCR products were cloned into the pGEM-T easy vector via TA cloning. As the PCR products had been generated using the high fidelity Roche Expand

Long polymerase, it was first necessary to add A nucleotide overhangs to the PCR product to enable cloning into pGEM-T. Briefly, 4µl of purified PCR product (75ng/µl) was incubated with 1µl 5x GoTaq reaction buffer, 1µl 2mM dATP and 0.2µl GoTaq polymerase. This mix was incubated in a PCR machine at 72°C for 30 min.

Following addition of the A overhangs, the PCR product underwent a ligation reaction with the pGEM-T vector. The ligation was carried out at a ratio of 1:3 vector to insert. Here the vector is 3kb in size and the insert is 700bp. Using the Promega Biomath calculator a 1:1 ratio would require 11.7ng insert to 50ng vector. Therefore $11.7 \times 3 = 35$ ng insert required for ligation. The ligation was then carried out according to the manufacturers recommended protocol. Briefly, the ligation mix contained 5µl 2x ligation buffer, 1µl pGEM-T vector (50ng), 0.7µl PCR insert, 1 µl T4 DNA ligase and 2.3µl water up to a final volume of 10µl. The ligation was mixed by pipetting and left overnight at 4°C.

2.2.3.4 Mutagenesis

Mutagenesis was performed for one exonic deletion variant (del Ex32-34), on order to generate this deletion in the pCI-neo vector. This was to ensure that all deletion mutants were then in identical expression vectors, enabling a direct comparison of all mutants with WT.

Site directed mutagenesis was performed using the QuikChange Lightning kit (Agilent). Mutagenesis reactions were set up according to the Agilent QuikChange protocol. Briefly, 125ng each of the forward and reverse mutagenesis primers were added to a tube containing 5µl of 10x reaction buffer, 1µl VWF-pCIneo (100ng/µl), 1µl dNTPs, 1.5µl QuikSolution and made up to a final volume of 50µl for each reaction. 1µl of QuikChange Lightning Enzyme was then added to each tube before running the PCR protocol shown in Table 3.4 Following the mutagenesis PCR, 1µl of *Dpn I* enzyme was added and samples were incubated at 37°C for 1 hour to degrade the parental DNA.

2.2.3.5 Transformation of competent cells

Transformations were performed in NM554 cells following cloning and XL10-Gold cells following mutagenesis. The methods for each are different.

For NM554 transformations, 2 µl of ligation product/positive control DNA/vector only was added to 50 µl NM554 cells in sterile 1.5 ml tubes. The cells were gently mixed by flicking the tube. Transformations were incubated on ice for 20 min prior to heatshock for 45s at 42°C. The tubes were then immediately placed back on ice for 2 min prior to the addition of 950µl of room temperature SOC media. Tubes were then incubated at 37°C with shaking at 200 RPM before spreading 100 µl of each transformation onto LB agar plates containing ampicillin.

XL10-Gold competent cells were gently thawed on ice and 45 µl was added to a pre-chilled 14 ml round-bottom tube. 2 µl *Dpn I* digested DNA from the mutagenesis reaction was then added, samples were gently mixed and incubated on ice for 30 min. Transformation reactions were then heat pulsed at 42°C for 30 s before immediately returning to ice for 2 min. Next 0.5 ml of pre-warmed NZY⁺ broth was added to each tube and samples were incubated at 37°C for 1 hr with shaking at 200 RPM. Finally

a range of volumes (50-200 µl) of the transformations were then plated on LB agar plates, containing 50µg/µL ampicillin and incubated overnight at 37°C.

2.2.3.6 Growth and storage of E.coli used for plasmid preparations

Competent cells transformed with either WT or the desired mutant VWF plasmid construct were stored in glycerol stocks at -80°C. Glycerol stocks consisted of 25% glycerol (v/v) in LB broth containing the transformed E.coli from an overnight growth period at 37°C.

2.2.3.7 Plasmid DNA purification (Miniprep)

Following transformation of competent *E.coli*, multiple colonies were picked and used to inoculate separate LB broth cultures containing ampicillin prior to small scale plasmid purification using the QIAprep spin miniprep kit.

Colonies were picked and used to inoculate 3 ml LB broth containing ampicillin in a 15 ml round bottom tube. Cultures were incubated at 37°C overnight in an orbital shaker at 200 RPM. Following incubation, 1 ml of overnight bacterial suspension was added to a 1.5 ml microfuge, centrifuged at 5000 g to pellet the cells and the supernatant was discarded. This process was repeated with an additional 1 ml from the same culture in order to concentrate the bacterial pellet. The bacterial cell pellet was then resuspended in 250 µl cold buffer P1, ensuring no cell clumps remained. 250 µl of buffer P2 was then added and the suspension was incubated at room temperature for 5 min to lyse the cells. Following lysis, the reaction was stopped with the addition of buffer N3, resulting in a cell debris and protein precipitate. Samples were then centrifuged at 13000 g for 10 min to pellet the precipitate. The supernatant, containing the plasmid DNA was then transferred to a QIAprep spin column and centrifuged at 13000 g for 1 min to bind DNA to the column. The flow through was discarded and the column was washed with 750 µl buffer PE and centrifuged at 13000 g for 1 min. After discarding the flow through, the column was centrifuged again to remove any wash buffer residue. The QIAprep column was then placed into a clean 1.5 ml microfuge tube and the DNA was eluted by adding 30-50µl dH₂O, incubating at room temperature for 1 min and then centrifuging at 13000 g for 1 min. This small scale plasmid DNA purification was used to send samples of plasmid DNA for sequencing to identify if the mutation had been introduced.

2.2.3.8 Plasmid DNA purification (Maxiprep)

Large scale purification of plasmid DNA was performed using the QIAGEN Maxiprep kit. This method was used to produce larger yields of plasmid DNA for downstream use such as sequencing across the entire *VWF* cDNA insert with multiple internal primers or for generating sufficient plasmid DNA containing the desired mutation so as multiple transfection experiments could be performed. The methodology is similar to that of the miniprep.

Briefly, a 250 ml culture of LB broth was inoculated using either a colony or a scrape from a previously stored glycerol stock. The 250 ml culture was grown overnight at 37°C in an orbital shaker. 250ml overnight LB cultures were

harvested by centrifuging at 6000 g for 15 min at 4° C. The supernatant was discarded and the bacterial pellet was resuspended in 10 ml of chilled buffer P1, ensuring no cell clumps remained. 10 ml buffer P2 was then added to the suspension and mixed thoroughly by inverting the tube multiple times in order to lyse the cells. The lysis reaction was incubated at room temperature for 5 min. The lysis reaction was then stopped with the addition of 10 ml chilled buffer P3 and mixed thoroughly by inverting, leading to the formation of a white precipitate containing cell debris and protein. The lysate containing the precipitate was then poured into a QIA filter cartridge and incubated at room temperature for 10 min, before gently inserting the plunger and filtering the lysate into a clean 50ml Falcon tube. 2.5mL Buffer ER was added to the filtered lysate to remove any endotoxins within the solution and the solution was incubated on ice for 30 min.

A QIAGEN-tip 500 was then equilibrated with 10ml buffer QBT, before the filtered lysate was applied and allowed to flow through. The flow through was discarded and the tip was washed twice with 30 ml of buffer QC to remove any contaminants. The DNA was then eluted with 15 ml buffer QN into a 50 ml Beckman centrifuge tube. The DNA was then precipitated by adding 10.5ml isopropanol at room temperature and mixing well by inversion. The precipitated DNA was then pelleted by centrifugation at 15000 g for 30 min at 4°C. The supernatant was then carefully discarded and the DNA pellet was washed with 5 ml of ice cold 70% ethanol before repeating the previous centrifugation step. Following the final centrifugation, the ethanol was carefully discarded and residual ethanol was removed using a Gilson P200 pipette. Finally, the DNA pellet was air dried for ~10-15 min before resuspension in 500 µl dH₂O.

2.2.3.9 Quantification of plasmid DNA

Plasmid DNA samples purified via miniprep and maxiprep protocols was performed spectrophotometrically using the Nanodrop ND1000 as described for patient DNA in 2.2.2.

2.2.4 Mammalian cell culture and *in vitro* expression

2.2.4.1 HEK293 cell culture

HEK293 cells were maintained in DMEM with glutamax and 10 % FBS (v/v) at 37°C in a 5% CO₂ atmosphere. Cells were passaged when they reached ~80-90% confluency, usually around twice per week. Briefly, growth media was removed from the T75 flask and the cell monolayer was washed once with 5 ml PBS to remove residual serum. 2 ml Trypsin-EDTA was then added and the cells were incubated at 37°C for 5 min, before firmly tapping the flask in order to detach the cells from the culture vessel. The trypsin was then neutralised by addition of ~8 ml DMEM and the cell suspension was pipetted up and down to ensure an even cell suspension. The suspension was then either split into fresh culture flasks or used for cell counting and downstream seeding and transfection experiments.

2.2.4.2 HUVEC cell culture

HUVECs were cultured in endothelial cell culture medium for a short period prior to counting and seeding onto cover slips at a seeding density of 5×10^5 per well containing a 22 mm² coverslip. HUVECs were fixed and stained for immunofluorescent microscopy 24 h after seeding. HUVECs were not normally maintained beyond 2 passages.

2.2.4.3 Freezing and long-term storage of HEK293 cells

To preserve batches of cells at earlier passages, HEK293 cells were stored in freezing medium consisting of 70% DMEM, 20% FBS and 10% DMSO, in cryoprotective vials either at 80°C or in liquid nitrogen for long term storage. Prior to freezing, cells were grown to ~90% confluency. The day before freezing the growth media was replaced.

HEK293 cells to be removed from cryo storage were thawed rapidly at 37°C. The freeze medium-cell suspension was then added to 10 ml of pre-warmed DMEM culture media with FBS and cells were incubated at 37°C overnight. Media was then replaced and cells were passaged as normal.

2.2.4.4 Cell counting and seeding

Prior to seeding, cell suspensions were quantified using a haemocytometer. Following the generation of a cell suspension, described above, 20 μl was pipetted onto a modified Fuchs Rosenthal haemocytometer counting chamber under a glass coverslip. The haemocytometer was then viewed under a light microscope with 10x objective. Cells were then counted from within one 1mm^2 square. Cells were also included in the count if they overlapped the top and left hand side of the central square grid-lines. Those cells overlapping the right hand side and bottom grid line of the central square were ignored. Figure 2.2 gives an overview of the layout of the haemocytometer counting grid. The cell count was equivalent to the number of cells in 0.1 μl of the cell suspension. Multiplying this cell count value by 10 therefore gave the number of cells per μl . The required number of cells was then divided by this value to give the volume of cell suspension required.

For microscopy experiments, HEK293 cells were seeded into 6-well plates containing 22x22 glass coverslip at a density of 1.5 to 2 $\times 10^5$ to avoid over confluency on the coverslip which could impair downstream imaging. For ELISA experiments, where no coverslip was used, cells were seeded at a density of 2.5 $\times 10^5$ per well.

As HUVECs were not transfected and only cultured on coverslips over a 24 h period, they were seeded at a higher cell density of $\sim 5 \times 10^5$ per well.

2.2.4.5 Transfections

HEK293 cells were transfected with either WT pCI-neo-VWF, mutant VWF or both in co-transfections, using either Lipofectamine LTX reagent or Xfect reagent according to the manufacturer's recommendations with some modifications. Briefly, for Lipofectamine transfections, 750 ng plasmid DNA was diluted in 250 μl DMEM without FBS and mixed, prior to the addition of 2 μl LTX reagent. Samples were mixed by flicking the tube and briefly centrifuged to collect the contents. Samples were incubated at room temperature for 15 min to allow for the formation of LTX-DNA lipocomplexes. DNA-lipocomplex mixtures were then

added drop-wise to cells at ~70% confluency, while gently agitating the plate. For Xfect transfections, cells were grown to ~50% confluency. Xfect transfection mixtures consisted of 750 ng plasmid DNA diluted in 100 µl Xfect buffer. Samples were mixed by vortexing for 5 s before adding 1.5 µl of Xfect polymer. Samples were mixed again by vortexing and then incubated at room temperature for 15 min. DNA-polymer complexes were added to the cells drop-wise with agitation.

Growth media was replaced after 24 h and the cells were maintained at 37°C in 5% CO₂ for a further 48 h, that is 72 h post transfection, prior to analysis.

2.2.4.6 Cell lysis

Following transfections for ELISA experiments, supernatants were harvested for analysis and cells were lysed. Cell lysis involved the removal of culture media, then the cells were washed twice with ice cold PBS. Next 1x passive lysis buffer containing a protease inhibitor cocktail was added to the cells. Cells were then incubated at 4°C with agitation to until fully lysed.

2.2.5 ELISA

The enzyme-linked immunosorbent assay (ELISA) is an immunological assay which utilises the high specificity of antibodies to bind an epitope on a target protein of interest. In this way it can be used to quantify the amount of the target protein within a sample. Primary capture antibodies are attached to a solid support such as a microtitre plate well. Non-specific binding sites are blocked with bovine serum albumin (BSA) before the sample mixture is added. The protein of interest is then bound by the capture antibody and it is detected and thus quantified via the addition of a second detection antibody that is conjugated to the enzyme horseradish peroxidase (HRP). Addition of the chromogenic substrate o-Phenylenediamine (OPD) then enables colorimetric determination of protein concentration via its oxidation to 2,3-Diaminophenazine in the presence of a peroxidase, which is orange-brown in colour and can be read in a plate reader at 490nm. The amount of coloured product is proportional to the amount of target protein in that samples. An overview of the principle of the sandwich ELISA is shown in figure 2.3. VWF:Ag from transient transfection experiments was

analysed via enzyme-linked immunosorbent assay (ELISA), using the paired match VWF antibody set and Visualize buffer kit reagents according to the manufacturers recommendations.

Briefly, flat-bottom, 96-well microtiter plates were coated overnight with polyclonal, anti-VWF capture antibody, diluted 1:100 in coating buffer. The next day, plates were brought to room temperature and the primary antibody coating mix was discarded. Wells were then blocked with 150 μ l blocking buffer (1% BSA in PBS) for 1 hour. After blocking, the wells were washed 3 times with wash buffer (0.1% v/v PBS/Tween 20). 100 μ l cell lysate and supernatant samples were then added to individual wells and the plates were incubated for 90 min at room temperature in a sealed wet-box to avoid drying out. Samples were then discarded and the wash step was repeated as above. The secondary, HRP-conjugated capture antibody, was then diluted 1:100 in sample diluent and 100 μ l was added to each well. Plates were incubated as before for 90 minutes. Following the final incubation, plates were washed as described, before adding 100 μ l of OPD substrate solution (25 ml containing 1 OPD tablet and 50 μ l of 6% hydrogen peroxide [H_2O_2]). Samples were incubated at room temperature for ~ 10 minutes to allow time for the OPD-HRP reaction to produce the desired colour change. Reactions were then stopped with the addition of 50 μ l 2M sulphuric acid (H_2SO_4) to each well. Plates were then analysed in a Varioskan plate reader at 490nm. Three repeat readings were recorded for each well and averaged to give the mean well value. A standard curve was prepared on the same plate for each experiment using a standard calibrator plasma in a serial dilution. Dilutions for the standard curve were made using sample diluent at 1/10, 1/20, 1/40, 1/80, 1/160, 1/320, 1/640, 1/1280, 1/2560, 1/5120, 1/10240 and a sample diluent only well.

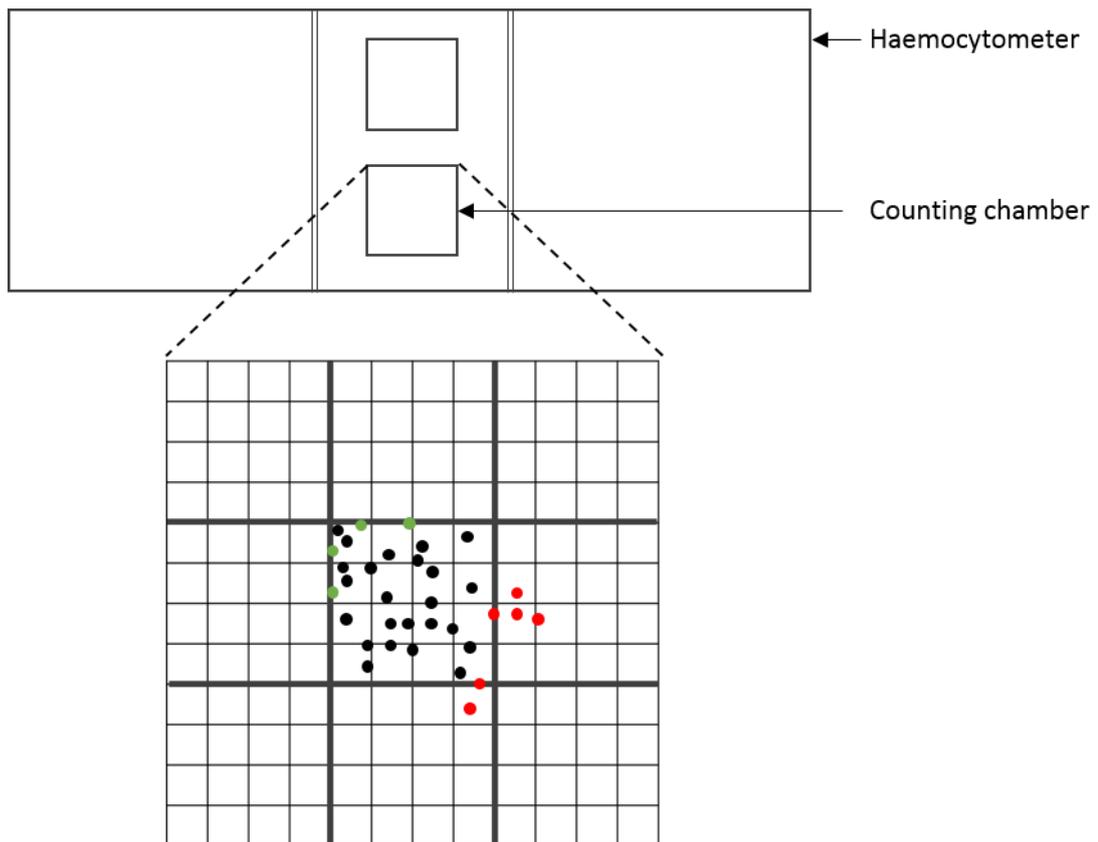


Figure 2.2 Cell counting using a haemocytometer

The haemocytometer grid consists of 12x12 small squares, making up a 3x3 grid of larger squares. Each of these 9 larger squares corresponds to a volume of 0.1 μl . It is the cells within one of these squares that was counted in order to determine the number of cells. Routinely, the central large square was used for counting. Those cells on the boundary grid line are counted if they overlap on the top or left hand side (green cells), whereas those that overlap the right hand and bottom grid line of the central square are not counted (red cells).

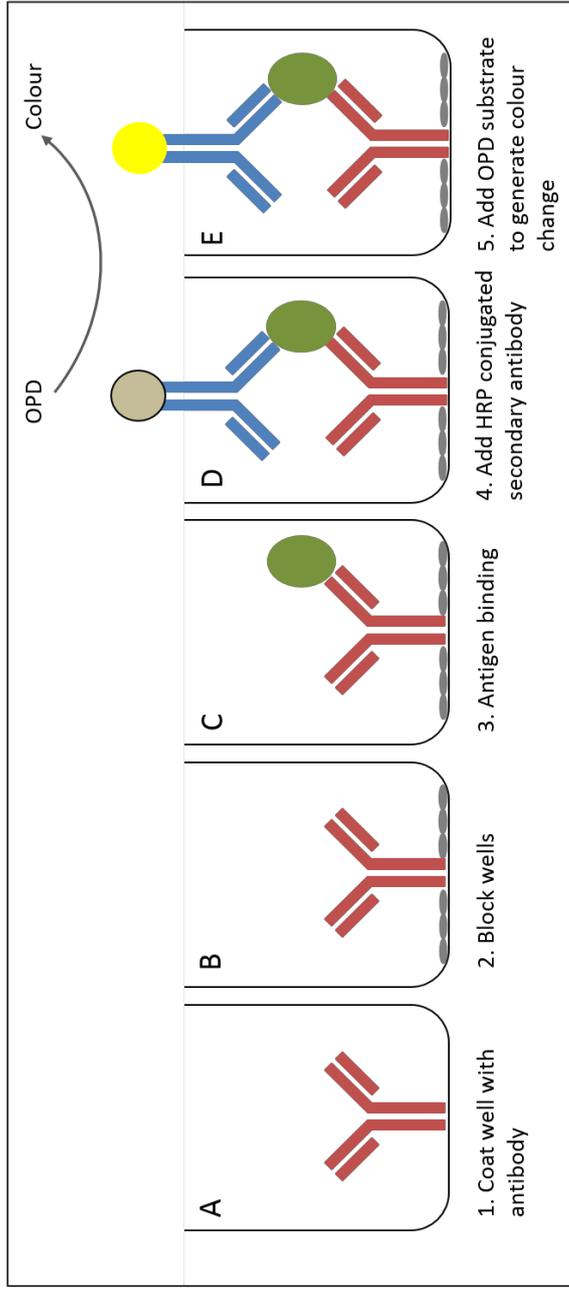


Figure 2.3 ELISA overview.

A. Capture antibody is coated on the wells of the 96-well plate. **B.** Excess capture antibody is removed and the wells are blocked with BSA blocking buffer. **C.** After washing to remove unbound BSA, cell lysates and/or supernatants containing VWF are added to the wells and are bound by the capture antibody. **D.** Further washing removes unbound residual antigen and then a secondary HRP conjugated detection antibody is added which binds the immobilised VWF antigen. **E.** Unbound secondary antibody is removed by washing and then the OPD substrate is added. In the presence of HRP and H_2O_2 , this is converted to a coloured DAP product, resulting in a quantifiable colour change. The reaction is stopped after ~ 10 min via the addition of H_2SO_4 and the colour change is measured via optical density at 490 nm in a Varioskan Flash plate reader.

2.2.5.1 ELISA data analysis

The standard curve results of known concentrations versus absorbance were plotted as a 4-parameter logistic standard curve. Using the 'interpolate unknowns from a standard curve' function in GraphPad Prism 6, absorbance values for supernatants and lysates were interpolated and transformed (via $X=10^X$) to give VWF concentrations. These VWF concentrations were then used for further calculations total VWF:Ag and secretion ratios.

Total VWF:Ag (lysate plus supernatant) was calculated for WT and each mutant and expressed as a percentage of WT total (WT = 100%). Each repeat experiment was normalised to the WT of that particular experimental repeat.

The ratio of secreted VWF:Ag to retained (lysate) VWF:Ag was given by VWF in supernatant/VWF in lysate. Ratios were then normalised to WT by dividing by the WT ratio. The WT ratio was then given to be 1.

Because the effective reduction in VWF secretion in the mutants is a function of the total VWF:Ag and the secretion ratio, then it follows that the effective VWF:Ag secretion was determined by multiplying the normalised secretion ratio by the total VWF (% of WT) i.e. WT = 1 x 100%.

2.2.6 Immunocytochemistry and fluorescent microscopy

2.2.6.1 Immunocytochemistry

Cells were fixed with BD Cytofix solution (BD) at room temperature for 20 min. Fixed cells were then washed three times with PBS and once with 50mM NH₄Cl to quench excess formaldehyde, before permeabilising with 0.2% (v/v) Triton X-100 in PBS. After blocking with 5% (v/v) normal goat serum in PBS, cells were washed 3x with PBS and then incubated with primary antibodies at the following concentrations: polyclonal rabbit anti-human VWF (1:1000), mouse anti-human calnexin (1:500) and sheep anti-human TGN46 (1:250). Cells were then washed 3x with PBS blocking solution and then 2x with PBS. Secondary antibodies described in section 2.1.6.1 were all used at 1:500 and incubated at room temperature for 1h. Following antibody incubations, cells were washed three times with 5% goat serum in PBS followed by 3 washes with PBS and 1 wash

with ddH₂O. Excess moisture was removed from the cover slips before mounting with Prolong Gold mounting medium.

2.2.6.2 Widefield fluorescent microscopy

Widefield images were captured on a Nikon TiE inverted microscope with a 100x oil objective and Andor Zyla sCMOS camera and saved as ND2 files. ND2 image files were opened and processed in FIJI (Schindelin et al., 2012) to make minor adjustments to contrast and brightness and were then saved as TIFF files. Images were presented as a maximum intensity projection of multiple z-stacks. The number of z-stacks varied according to sample thickness (~20-40 separate images).

2.2.6.3 Structured Illumination Microscopy (SIM) and 3D visualisation

Structured Illumination Microscopy (SIM) was carried out using a Deltavision V4 OMX 3D-SIM system fitted with a Blaze module (Applied Precision, GE Healthcare, Issaquah, USA). The objective was a 60 x 1.42 oil planapochromat lens, the system used a standard BGR filter set and used front illuminated sCMOS cameras, 2560 x 2160 pixels, pixel size, 6.45 μ m, readout speed, 95 MHz, 286 Mhz which were operated at 512 x 512, 15 bit for 3D-SIM imaging. SIM samples were illuminated using patterned laser illumination at 488nm and 568nm. For each z slice, samples were imaged in 5 phase shifts and 3 angles, z-steps were 0.125 nm. Reconstruction and alignment of the subsequent Moiré fringe raw image data was performed in SoftWoRx v6.5.2 (GE Healthcare) using optical transfer functions (OTFs) optimised for the specific wavelength and oil used. 3D reconstructed files were loaded into Imaris version 7.4 (Bitplane, Oxford, UK), to generate 3D movie files and apply surfaces to enhance visualisation of VWF positive structures. Surfaces were added using background subtraction and default settings. All raw images were processed in FIJI (Schindelin et al., 2012) and 3D reconstructions performed using Imaris v7.4.

2.2.7 Statistical analysis

ELISA data is presented as the mean and SEM of three independent experiments performed in triplicate. Data were analysed in GraphPad Prism version 6 (GraphPad Software Inc., San Diego, CA) using one-way ANOVA, assuming Gaussian distribution. Comparisons to WT were performed with a post-test Dunnett's multiple comparison analysis. Significance was defined as $p = < 0.05$.

3 Design, validation and application of a custom microarray for the detection of CNV in *VWF*

3.1 Introduction

Copy number variation (CNV) is the collective term used to describe quantitative structural genomic variations that include deletions and duplications (Conrad et al., 2010). These are distinct from other balanced structural variants such as inversions and translocations as they are unbalanced, resulting in a change in genomic content or copy number at a particular locus. The first reports on the incidence of CNV in the human genome came in 2004 when two independent studies reported on the contribution of large-scale CNV to genetic variation amongst individuals (Lafrate et al., 2004, Sebat et al., 2004). These findings highlighted the dynamic nature of the human genome and implicated CNV as a contributing factor to phenotypic variability between individuals and susceptibility to disease. Later studies led to the first comprehensive CNV map of human genome, based on samples from the HapMap consortium (Redon et al., 2006). Historically CNV were described as genomic imbalances hundreds of kilobases (Kb) in length, however, this was primarily due to the limitations and resolution of the detection methods used at the time rather than a specific size classification. Therefore, as surveys of CNV have been conducted at increasingly higher resolution, so smaller CNVs have been detected, thus redefining what is understood by the term. Results from these studies are incorporated into online databases such as the database of genomic variants (DGV), which currently describes CNV to be within the range of >50 bp to 3 Mb (MacDonald et al., 2014). Smaller insertion/deletions (indels) may be considered to be variants <50-100bp, that is those that are detectable in a single NGS read (Carvalho and Lupski, 2016). CNV and smaller insertion/deletions (indels) account for significantly more variation across the human genome than do SNPs (Pang et al., 2010) and are now considered to be a major driver in human evolution, health and disease (Zhang et al., 2009). The most up to date CNV maps suggest that around 5 – 10% of the human genome contributes to CNV in healthy individuals (Zarrei et al., 2015). However, while these are a useful reference for assessing the biological and clinical significance of variants, it is important to consider the limitations of CNV maps produced from databases such as DGV. Firstly, although healthy individuals are selected for these databases, detailed phenotypic data is often absent or insufficient for all disease phenotypes. For example, healthy controls used in the collation of these data are unlikely to have had any detailed

haematological analyses such as VWF:Ag levels etc. Disease is also a dynamic process, individuals may present with symptoms at different stages, despite being healthy at the time of analysis, and this would certainly be an issue with diseases associated with late onset. Finally some diseases may go undiagnosed and so associations of a particular CNV with a disease phenotype is not possible.

3.1.1 The evolutionary origin of CNV. Selective advantages and medical relevance

The focus of CNV analysis has predominantly been on its role in human disease, particularly microdeletions and duplications associated with congenital abnormalities and developmental delay. Examples of CNV that are the result of adaptive evolution and therefore confer a selective advantage are relatively few. Notable examples include CNV in the gene encoding amylase (AMY), with increased copy number being associated with increased salivary amylase and predominating in cultures/populations in which high dietary starch intake is a feature (Perry et al., 2007). Segmental duplications of the *CCL3L1* gene confer some protective effect from HIV/AIDS, with individuals that have a lower copy number being more susceptible to disease (Gonzalez et al., 2005). More recently, exonic deletions in the haptoglobin gene have been associated with reduced blood cholesterol levels, therefore suggesting a beneficial effect of these mutations (Boettger et al., 2016).

3.1.2 CNV in VWD

The earliest report of CNV in VWD came in 1987, with the identification of large deletions in type 3 patients who had also developed inhibitory alloantibodies to VWF (Shelton-Inloes et al., 1987). Later, heterozygous and homozygous whole gene deletions were identified in type 3 families via Southern blot analysis (Ngo et al., 1988). As our general understanding and appreciation of the role of CNV in human disease has developed, pathogenic CNV in *VWF* as a cause of VWD has become well established. Numerous studies to date have described deletion and duplication variants involving either single exons, multiple exons or whole gene deletions of the *VWF* locus. These *VWF* CNV range in size from ~2 Kb to

over 250 Kb for the whole gene deletions, which in some cases extend into the flanking gene *ANO2* (Schneppenheim et al., 2007). A summary of *VWF* CNV is provided in table 3.1. It is clear from this overview that the prevalence of duplications in the *VWF* locus is significantly lower than deletions with only two previous reports including a duplication of exon 6 (Boisseau et al., 2013) and a duplication of exons 9-10 (Obser et al., 2016).

CNV have been associated with all three types of VWD, however the majority of initial findings stem from studies on type 3 patients with homozygous deletions that were detectable via standard sequencing approaches. For example, the identification of the exon 4-5 deletion by Sutherland *et al* was facilitated by the initial identification of a homozygous deletion in type 3 patients and the subsequent development of a deletion-specific multiplex PCR assay and analysis of patient cDNA (Sutherland et al., 2009). Without the homozygous type 3 deletion, it is likely that the heterozygous deletions detected in type 3 compound heterozygotes and type 1 patients would have remained elusive, due to the masking effect of the normal allele which is amplified.

This masking effect is likely to have been an issue in many previous genetic analyses of VWD cohorts that have failed to identify CNV. For example, the large EU, UK and Canadian type 1 VWD cohort studies failed to identify candidate mutations in ~30-35% of cases (Cumming et al., 2006, Goodeve et al., 2007, James et al., 2007). Therefore the prevalence of pathogenic CNV in *VWF* is likely underestimated in the VWD population. Key changes in this respect have been brought about by the development of molecular tools able to detect copy number changes.

Table 3.1 Summary of CNV reported in *VWF*

Study	CNV(s) detected (kb)	Exon(s) affected	VWD type
(Peake et al., 1990)	Del	Ex42	3
(Bernardi et al., 1993)	Del	Ex26-34	2
(Mancuso et al., 1994)	Del	Ex22-43 Ex33-38	3 3
(Schneppenheim et al., 1994)	Del Del	Ex14-52 Whole gene	3 3
(Baronciani et al., 2003)	Del	Ex23-52	3
(Xie et al., 2006)	Del (61 Kb)	Ex6-16	3
(Schneppenheim et al., 2007)	Del (253 Kb)	Whole gene	3
(Mohl et al., 2008)	Del	Ex1-3	3
(Sutherland et al., 2009)	Del (8.6 Kb)	Ex4-5	1 and 3
(Bellissimo et al., 2011)	Del	Ex1-3 Ex4-5 Ex18 Ex35-38	3 3 3 3
(Cabrera et al., 2011)	Del (84 Kb)	Ex16-43	3
(Johansson et al., 2011)	Del	Ex16-52	1
(Nik-Zainal et al., 2011)	Microduplication Ring chr12	Ex35-52	none
(Yadegari et al., 2011)	Del Del Del Del Del	Ex1-5 Ex4-34 Ex6 Ex33-34 Ex48-52	3 3 3 2 3
(Solimando et al., 2012)	Del Del	Ex1-3 Ex17	3 3
(Ahmad et al., 2013)	Del (2.2 – 80.8 Kb)	Ex14-15	3
(Boisseau et al., 2013)	Dup Del Del	Ex6 Ex6-18 Ex18	1/3 3 3

	Del Del Del Del	Ex19-20 Ex32-34 Ex33-34 Ex37-38	1
(Bowman et al., 2013)	Del Del	Ex4-5 Ex19-20	3 3
(Cartwright et al., 2013)	Del (2 Kb) Del (5.2 Kb) Del	Ex3 Ex32-34 Ex33-34	1 1 1
(Hampshire et al., 2013)	Del (9.7 Kb)	Ex17-18	3
(Theophilus et al., 2013)	Del Del Indel? Del	Ex4-52 Ex19-52 Ex22 Ex33-34	3 3 1 2
(Kasatkar et al., 2014)	Del	Ex4-49 Ex11-16 Ex16-52	3 3 3
(Daidone et al., 2015)	Del	Ex32-34	1
(Obser et al., 2016)	Dup (5 Kb)	Ex9-10	2A

3.1.3 Detecting CNV

3.1.3.1 MLPA

Developed by Schouten and colleagues in 2002, multiplex ligation dependent probe amplification (MLPA), enables the detection of copy number variation of single exons (Schouten et al., 2002). This molecular technique has been utilised to identify copy number changes in a number of diseases including VWD.

Initial studies in VWD patients demonstrated the utility of MLPA in identifying heterozygous deletions that were previously undetected. Cabrera and colleagues were the first to utilise MLPA to identify a deletion of exons 16-43 in homozygous type 3 patients and also in their heterozygous relatives, in whom the deletion could not be identified via standard methods (Cabrera et al., 2011). Subsequently the use of MLPA was promoted as a means of screening for large deletions in type 3 patients for whom no second mutation had been detected (Yadegari et al., 2011, Bowman et al., 2013). This application of MLPA for detecting compound heterozygous deletions in type 3 patients found use in numerous other studies (Boisseau et al., 2013, Hampshire et al., 2013, Theophilus et al., 2013) as outlined in table 3.1. In addition to detecting compound heterozygous deletions in type 3 patients, MLPA has been successfully utilised to demonstrate that heterozygous deletions underlie a proportion of type 1 VWD cases (Cartwright et al., 2013, Theophilus et al., 2013).

Previous studies on type 1 VWD IC carried out by the Sheffield Haemostasis research group used MLPA to identify CNV in six type 1 IC, including three novel, in frame, heterozygous deletions of exon 3, exons 32-34 and exons 33-34 (Hampshire et al., 2010, Cartwright et al., 2013). Following these MLPA studies on 94 patients from the EU study for whom, no candidate mutation was identified or where the mutation did not explain the VWD phenotype, there remained a significant proportion of cases in which MLPA did not detect CNV. Similar observations were made by other collaborating laboratories in the USA (Blood Centre, Milwaukee) and Canada (Queens University, Kingston), resulting in a subset of VWD patients from all three VWD types, for which the genetic basis of the patient phenotype could not be explained.

One possible explanation for these findings relates to the inherent limitations of the MLPA technique. By design, MLPA is a targeted technique for assessing exonic CNV. It is limited to around 40 target sequences per reaction which makes it more suited to focusing on a specific investigation into the presence of exonic CNV at a particular genomic locus. A second limitation of this technique is the MLPA probes themselves. These are designed to sit in close proximity to each other in a particular exon, however they do not cover the entire exon, so smaller copy number aberrations within an exon and outside the MLPA probe annealing and ligation sites will not be detected. Furthermore, CNV that span introns and exons and perhaps only include a partial region of an exon may not be detected. Intronic and intergenic CNV of potentially important functional regulatory regions and splice sites are also undetectable via this approach. This final point is reflected in the current literature by the absence of any reports of non-coding CNV in *VWF*. With this in mind it is clear that alternative methods are desirable for a more comprehensive interrogation of the *VWF* locus for coding and non-coding CNV.

3.1.3.2 Array comparative genomic hybridisation

Array comparative genomic hybridisation (aCGH) is a widely used method for copy number analysis, the principle and methodology of which has been described in chapter 2. This technology is an evolution of earlier molecular cytogenetic techniques used in the diagnostic analysis of solid tumours, in which labelled DNA was hybridised to metaphase chromosomes (Kallioniemi et al., 1992). Taking the principle of comparative genomic hybridisation and applying it to a more robust system in the form of DNA probes immobilised on a glass slide simplified the procedure, enabled improved resolution and the DNA microarray was born (Solinas-Toldo et al., 1997, Pinkel et al., 1998). This development of array-based comparative genomic hybridisation has enabled the analysis of CNV across the entire genome and its role in numerous diseases to be explored (Pinkel and Albertson, 2005). Indeed it is now considered the gold standard by diagnostic laboratories for the detection of CNV and is widely used for the genetic diagnosis of developmental delay disorders and congenital abnormalities (Miller et al., 2010).

Custom microarrays consisting of probes specific to a particular gene or groups of genes allow higher resolution CNV analysis of a specific region of interest. Custom aCGH has been demonstrated successfully for Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD). Hegde and colleagues, developed a custom 385K microarray consisting of 45-60mer oligonucleotide probes enabling them to reassign breakpoints of known abnormal samples and identify 35 deletions and 14 duplications in new samples received for diagnostic testing (Hegde et al., 2008). More recently, custom microarray design has been used to confirm CNV and guide subsequent deletion breakpoint mapping experiments in rare diseases such as Usher Syndrome (Steele-Stallard et al., 2013) and the ciliopathy Bardet-Biedl syndrome (Lindstrand et al., 2016).

In the VWD field, custom aCGH has been applied to target the *VWF* locus in a group of fourteen type 3 patients that, following DNA sequence analysis, had no mutation in one or both alleles. Deletions were detected in four of these patients (29%) (Bellissimo et al., 2011). However, a major limitation of aCGH in the context of VWD analysis is the presence of the partial pseudogene. Due to the high level of sequence homology (~97%), differentiation between the *VWF* sequence and the pseudogene is problematic and this region was omitted from the custom array design used by Bellissimo and co-workers. This does not imply that aCGH would not detect CNV in this region, if included in a custom probe set, however, care would need to be taken with interpretation of results and any positive findings may require follow up and confirmation by PCR and sequencing. Despite this issue, it remains that aCGH offers great potential for CNV screening in VWD without the limitations of MLPA. Finally, another significant advantage of the aCGH approach is the ability to incorporate probes from across the entire human genome, potentially presenting opportunities for the identification of novel CNV in genetic loci outside *VWF*. In theory, this could enable the identification of other key genes that influence VWF biology.

3.2 Hypotheses

Based on previous findings and the current understanding of CNV prevalence in *VWF*, three experimental hypotheses were constructed:

- i. A proportion of CNVs within *VWF* would not have been detected using the MLPA approaches described above and these undetected CNV are a contributing factor to the VWD mutation spectrum.
- ii. Non-coding CNV (intronic and intergenic) causing disruption of functional elements such as promoters, transcription factor binding sites, enhancers and splice sites contribute to VWD.
- iii. Patients from all three VWD classifications for whom no prior mutation has been found or where identified mutations do not fully explain the phenotype are valid candidates for CNV screening.

3.2.1 Aims

In order to test the hypotheses put forward in 3.2, the aim of this study was to design, validate and utilise a custom *VWF* specific microarray that would enable a more comprehensive CNV screen of the entire *VWF* locus, including introns, exons and 3' and 5' intergenic flanking regions. This microarray would then be used to screen for CNV in a subset of VWD patients including those diagnosed with VWD types 1, 2 and 3. Ultimately, this would enable a clearer understanding of the frequency and contribution of coding and non-coding CNV, to the pathology of all three VWD types.

3.3 Results

3.3.1 Patient DNA quality assay

High quality, intact, high molecular weight DNA is essential for successful aCGH analysis, therefore prior to the initial validation experiments, assessment of the DNA quality was carried out using the DNAOK assay. These results indicated that the majority of samples all produced the correct number of amplicons from the multiplex PCR, suggesting high quality intact DNA (Figure 3.1).

3.3.2 Custom VWF array design: version 1

The custom probe set, consisting of 8 probe groups was generated in SureDesign and visualised in the UCSC genome browser to assess coverage and spacing (Figure 3.2) as described in chapter 2. A BED file consisting of ~700 probes for the *VWF* array described in 3.1.3.2, was kindly provided by Dan Bellissimo (University of Pittsburgh, USA). As described previously, this design lacked probe coverage in the pseudogene region and the 5'/3' intergenic flanking regions. In the current design, this was improved upon with the addition of 237 probes in the pseudogene region and 502 probes in the intergenic flanking regions. Remaining gaps in the probe coverage represent highly repetitive elements, lacking sufficient unique sequence to warrant probe coverage. The final array design contained ~1800 probes spanning both 5' and 3' intergenic regions (IR), up to *CD9* and *ANO2* respectively.

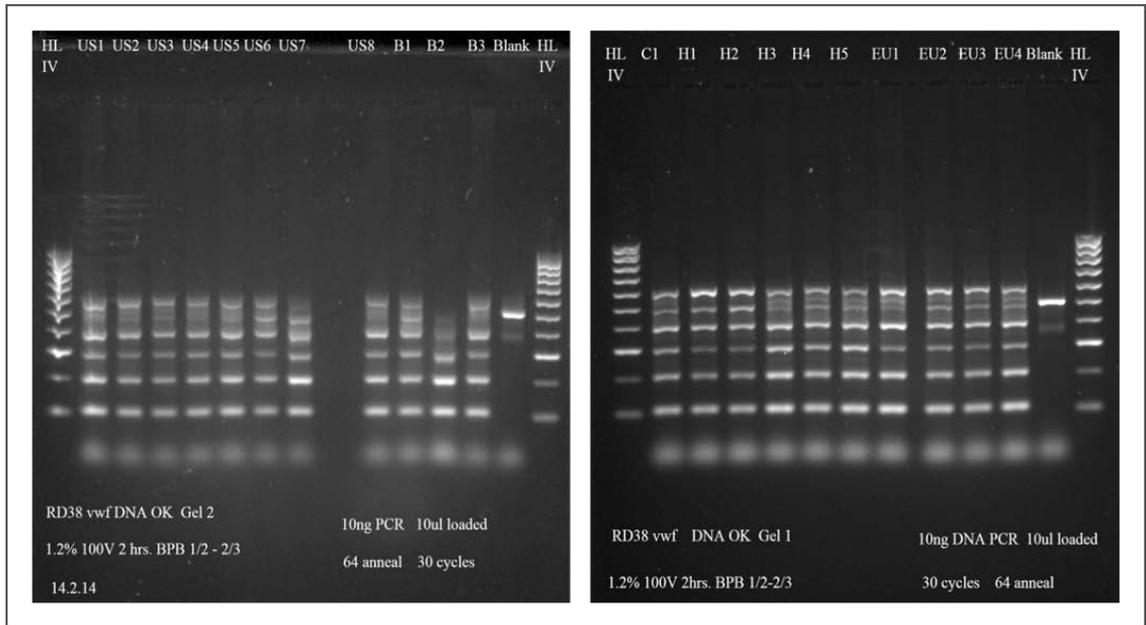


Figure 3.1 Example of DNAOK! quality assay results.

Good quality, intact DNA should produce 6 amplicons ranging from 600bp down to 100bp. Note that sample B2 (top left gel) shows partial degradation with the loss of the 500bp and 600bp bands. The blank lane contains an internal control and no patient DNA.

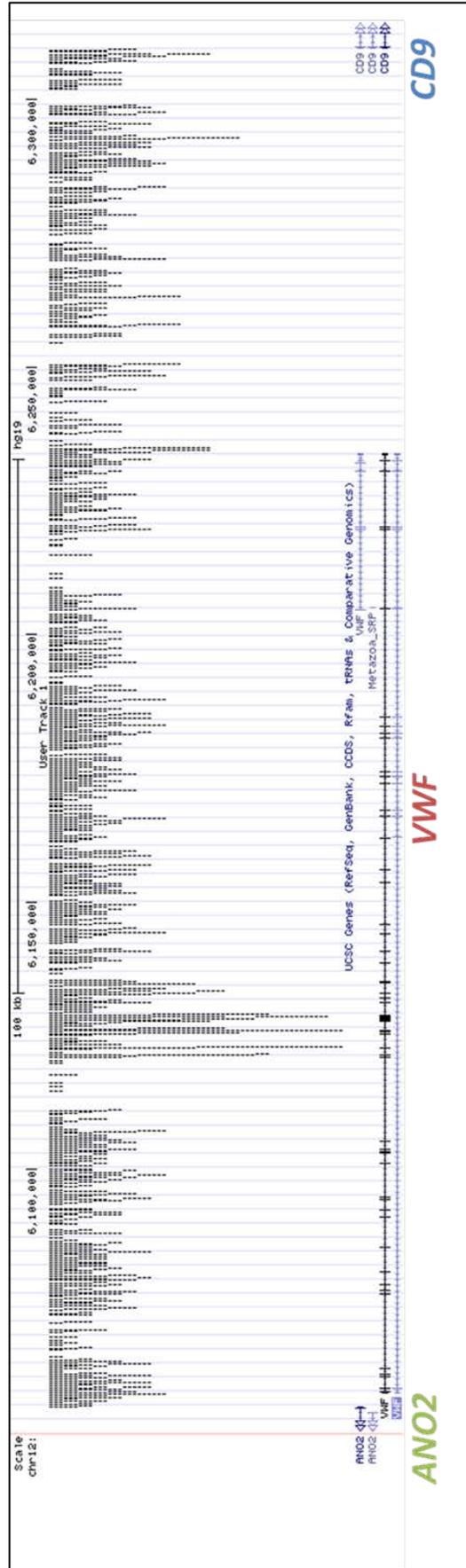


Figure 3.2 Probe coverage across VWF in custom array design. ~1800 probes in total were selected to span the VWF locus including introns and 5' and 3' UTRs. Coverage spans ~260kb of genomic DNA on chromosome 12p13.3. Larger gaps in probe coverage represent regions of highly repetitive elements in VWF.

3.3.3 Array Validation

3.3.3.1 Initial validation analysis of design version 1

In order to validate the performance of the custom array, 18 IC with known CNV in *VWF* were selected for initial analyses. These included patients from all three VWD types and provided a wide range of CNVs to test the array performance (Table 3.2). This preliminary data could then be used as a basis for further design modifications as necessary. Of these 18 IC, 12 had CNV with previously mapped breakpoint coordinates and 3 of these were within the pseudogene region. The remaining CNV samples with no mapped breakpoints involved 3 within the pseudogene region and 3 outside. All known CNV used for validation purposes were exonic variants.

Analysis of IC UK-M C48, carrying a previously reported heterozygous whole gene deletion (Schneppenheim et al., 2007), demonstrated a clear shift of probes from within the 5' intergenic region through to the end of exon 52 in *VWF*, towards a $-1 \log_2$ ratio, indicating a clear loss of one *VWF* allele (Figure 3.3). The sensitivity of the array was then assessed with analysis of IC UKD17. This patient carries a homozygous whole gene deletion of *VWF* which the array was able to detect as a copy number change of -2 , indicating the loss of both *VWF* alleles in this patient (Figure 3.4).

Table 3.2 List of index cases with previously identified CNV across the *VWF* locus used in the validation study.

Patient ID	Sheffield array ID	CNV	Breakpoints mapped	Size (kb)
1585PM	B1	del Ex33_ Ex34	No	?
2625 SN	B2	del Ex22	No	?
UK-M C48	H3	del Ex1_ Ex52 heterozygous	yes	253
UK D17	H1	del Ex1_ Ex52 homozygous	Yes	253
LM12567	H5	dup Ex9_ Ex10	yes	5
AU0020	US2	del Ex33_ Ex34 + ?dup Ex14_ Ex16	No	?
EU P6F1 III2	EU4	del Ex33_ Ex34	no	?
EU P9 F11 I1	EU1	del Ex3	Yes	2.1
SS D25	H4	del Ex7 + invEx8_ Ex13 delEx14_ Ex52	Yes Yes	411
MK0088	US5	del Ex4_ Ex5	Yes	8.6
PB0092	US8	del Ex35_ Ex38	No	?
8028TB	B3	del Ex4_ Ex52	No	?
IA0080	US1	del Ex18	Yes	3.8
PB0081	US7	del Ex4_ Ex5 del Ex1_ Ex3	Yes No	8.6
AU0101	US3	Possible del Ex33_ Ex34 ?mosaic	No	?
MMK0038	US4	small del Ex28	No	?
P9 F3 II1	EU3	del Ex32_ Ex34	Yes	5.2
FIII I2	T1	del Ex17_ Ex18	Yes	9.7

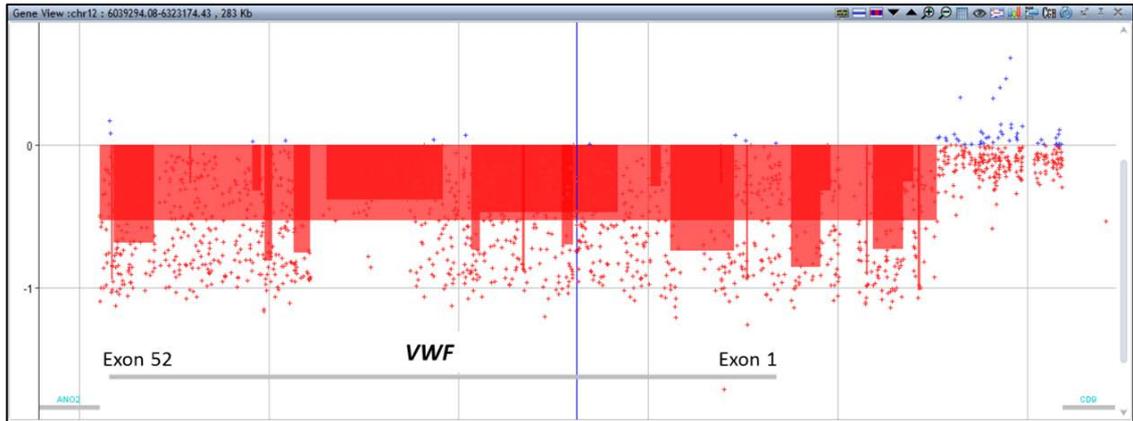


Figure 3.3 IC UK-M C48, heterozygous whole gene deletion of VWF. This patient has a -1 copy number reflecting the loss of one allele. The deletion extends into the 5' IR. The group of probes on the far right of the figure represent a normal copy number as this region is not deleted in this patient.

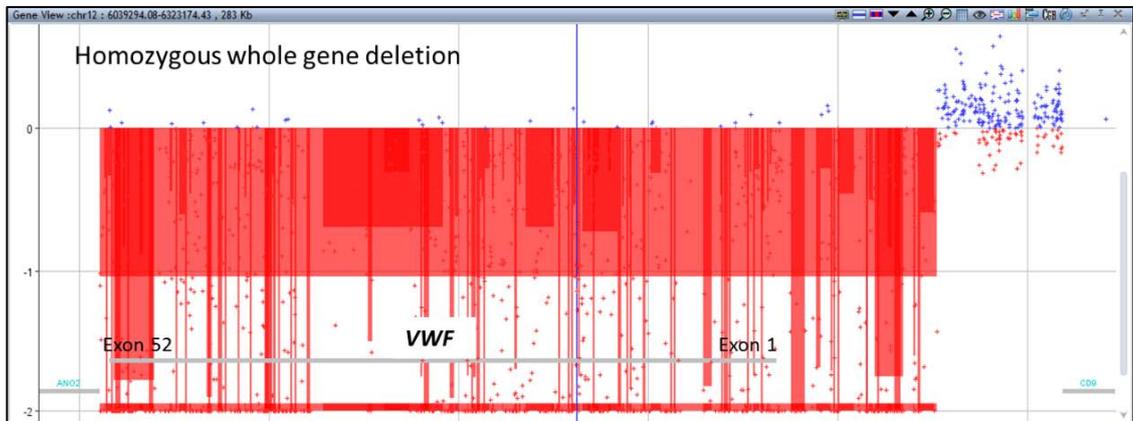


Figure 3.4 IC UK D17, homozygous whole gene deletion of VWF. This patient has a -2 copy number reflecting the loss of both alleles. The deletion extends into the 5' IR.

Having established that the array could detect an entire loss of the *VWF* locus and differentiate between heterozygous and homozygous states, the next stage was to establish if smaller, more subtle CNVs could be detected using this method. Starting at the 5' end of *VWF* with IC P9F11 I1, the array was able to detect a deletion of ~2 kb, however this was within intron 3 only and did not indicate a deletion of exon 3 (figure 3.5). According to the previously mapped data for this patient, the deletion is around 2kb in size (see table 3.2) and deletes the entire coding sequence of exon 3, suggesting that the probe coverage around this region was insufficient.

In IC PB0081, the array detected a previously identified deletion of exons 1 to 3. This deletion was ~35kb and extended from the 5' intergenic region into intron 3, similar to the deletion of a different patient mapped by Mohl and colleagues (Mohl et al., 2008). The array breakpoints mapped to within 469 bp (3') and 186 bp (5') of the identified breakpoints. This patient also carries a mapped 8.6kb deletion of exons 4 and 5. The array was able to detect this deletion and the software made a call of 9.3kb, indicating breakpoint positions of 1.3 kb off the mapped 3' breakpoint and 0.6kb off the mapped 5' breakpoint (array data not shown). The same exon 4-5 deletion was also present in a second IC, MK0088 and the array detected identical breakpoint positions, indicating the specificity of the probe performance.

In order to assess the efficacy of the array in detecting duplications within *VWF*, IC LM12567 with a known duplication of exons 9 and 10 was analysed. The array confirmed this imbalance and, taking into account the maximum breakpoints, indicated a duplication of 4.2 kb (Figure 3.6). This compared favourably with the mapped breakpoints which correspond to a 5 kb duplication (Table 3.2) and illustrated that dosage changes from 2 copies to 3 of *VWF* could be discerned.

To further assess the array performance over the wider *VWF* locus, two further IC were analysed; IA0080 carrying a deletion of exon 18 and FIII I2 carrying a deletion of exons 17-18. In the first of these patients the array detected a 3.3kb deletion of exon 18 when using the maximum breakpoints, a difference of 491bp at the 3' end and just 26bp off the 5' position of the mapped deletion. A similar level of breakpoint accuracy was observed for the exon 17-18 deletion, where the

array detected a deletion of 9.5 kb compared to the 9.7 kb of the mapped deletion. In this instance the array breakpoints were accurate to within 70bp of the mapped 5' and 3' breakpoints.

Table 3.3 Comparison of mapped breakpoints to aCGH breakpoints.

Patient ID	CNV	Size (bp)	Start breakpoint	End breakpoint	Array CNV size (kb) (maximum)	Array start (maximum)	Array end (maximum)	Array resolution (bp) 3' / 5'
UK-M C48	del Ex1_ Ex52 het	253247	6023118	6276365	220	6055441	6276288	32323* / 77
UK D17	del Ex1_ Ex52 homozygous	253247	6023118	6276365	220	6055441	6276288	32323* / 77
EU P9 F11 I1	del Ex3	2086	6228754	6230840	1.6	6228640	6230224	114 / 616
PB0081	del Ex1_ Ex3	35540	6226894	6262433	35.3	6227363	6262619	469 / 186
	del Ex4_ Ex5	8630	6212481	6221111	9.3	6211149	6220480	1332 / 631
MK0088	del Ex4_ Ex5	8630	6212481	6221111	9.3	6211149	6220480	1332 / 631
LM12567	dup Ex9_ Ex10	5047	6176741	6181788	4.2	6177532	6181772	791 / 66
IA0080	del Ex18	3780	6150646	6154426	3.3	6151137	6154452	491 / 26
FIII I2	del Ex17_ Ex18	9682	6149774	6159456	9.5	6149842	6159383	68 / 73
SSD25	del EX7	18000	6107519	6125464	Noisy data	Noisy data	Noisy data	
	14-52 inv 8-13	400000	5712029	6123835				

*Ex1_ Ex52 deletion extends into ANO2 where there is no probe coverage in the array design and thus a large discrepancy is seen in the 3' breakpoint coordinates. IC SSD25 is omitted from this list as it contains a complex rearrangement and the deletion extends significantly into ANO2 where the array has no probe coverage. Maximum breakpoints referred to in this section refer to the size of the CNV measured from the two flanking un-called probes at either of the CNV. Maximum sizes are often quoted in a diagnostic aCGH service as it is not possible to determine how far the CNV extends beyond the called start and stop breakpoint probes, up to the next probe that shows a normal copy number.

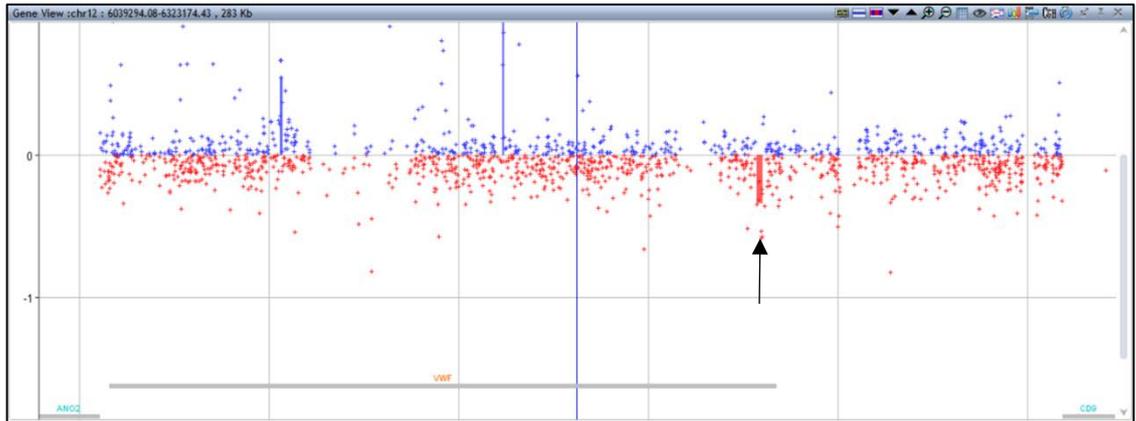


Figure 3.5 IC EU P9 F11 I1, 1.6kb deletion of exon 3. The smallest deletion analysed on the array as part of the validation. Black arrow indicates site of deletion.

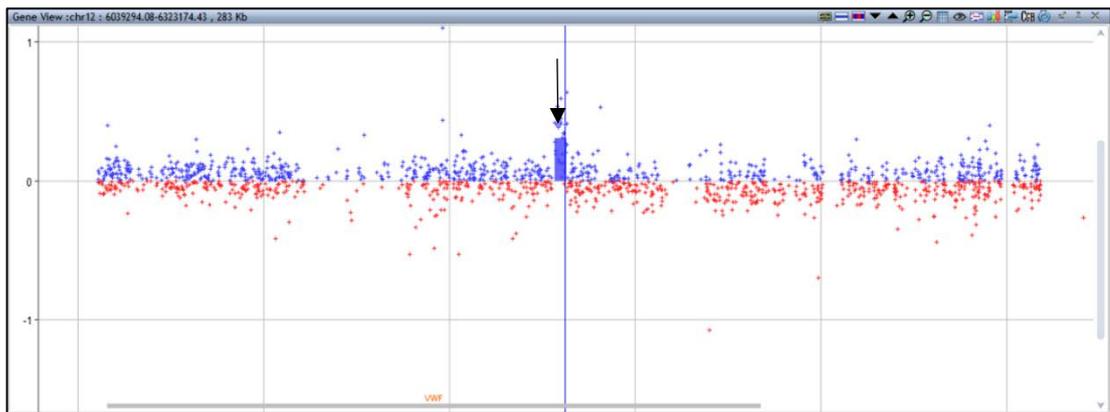


Figure 3.6 IC LM12567, duplication of exons 9 and 10. Black arrow indicates 3.3kb duplication

IC SSD25 was previously identified as having a complex rearrangement in *VWF* involving an inversion of a segment including exon 8 through to exon 13 resulting in the deletion of exon 7, and a large deletion of exons 14 to 52 which extends into *ANO2*. Data from this array was somewhat noisy and therefore data analysis was compromised. The array would not be able to detect a balanced inversion, however, deletions of intron 6 and exons 14 through to 52 were detected. Furthermore, with careful manual analysis, the intron 6 deletion expanded its 3' breakpoint from 6184903 downstream to 6183929, resulting in a 1.5kb deletion encompassing exon 7.

No previous breakpoint mapping data was available for IC 8028TB (del Ex4_Ex52), PB0092 (del Ex35_Ex38) and 2625SN (deletion of exon 22). Analysis of these samples therefore could not be used to assess the array resolution, but did suggest approximate breakpoints for these patients. These results are summarised in table 3.4. For IC 2625SN, no exon 22 deletion was detected.

Table 3.4 IC with no breakpoint data.

Patient ID	CNV	Array CNV size (kb) (maximum)	Array start (maximum)	Array end (maximum)	Confirmation of exon deletion
8028TB	del Ex4_ Ex52	167	6055645	6222918	del Int3_ Ex52
PB0092	del Ex35_ Ex38	9.35	6098419	6107769	del Ex35_ Ex38
2625SN	del Ex22	NA	NA	NA	No

Approximate breakpoints are shown based on array data.

3.3.3.2 Detailed manual probe calling is necessary for pseudogene CNV detection in design version 1

A total of 6 IC with CNV within the pseudogene region were analysed during the array validation and 3 of these (EU P9F3II1, EU P6F1 III2 and MK0038) had mapped breakpoint data available (Table 3.5). Initial analysis of these cases, using the default analysis method in Cytogenomics showed a poor detection rate of the known deletions summarised in Table 3.3, with the exception of 1585PM in which an 8.7 kb (max) deletion was automatically detected in intron 34 (figure 3.7). To improve the automated detection in this region a modification was made to the analysis method, switching off the design level filter and thus removing any similarity probe filters. This allowed more probes within the pseudogene region to be used in the copy number analysis, resulting in slight improvements to the automated detection rate in this region. This method was named Trial-method-1. This modified analysis method was combined with a manual detection approach whereby the maximum deletion size was determined from the two probes flanking the shifted region. Using this method, a 10kb deletion in IC 1585PM was detected with a 5' breakpoint in intron 33 and a 3' breakpoint in intron 34 (table 3.3), resulting in a deletion of exon 34.

Breakpoint mapping data for IC P9F3II1 indicated a ~5.2kb deletion of exons 32 – 34, while the array detected a deletion of ~3.8kb, deleting exons 32-33 only (Figure 3.8). However, closer analysis of the array probe coverage showed a 305bp gap between the last 3' probe shifted and the next flanking probe. The small, 178bp exon 34 sits within this region, therefore when maximum breakpoints were taken into account up to this next flanking probe (coordinates 6120861), 100 bp of exon 34 was incorporated in a larger 4.1kb deletion. Similar conclusions could be made for IC AU0020 and EU P6F1 III2, indeed all three of these patients showed the same 3' breakpoint coordinate of 6121116 (Table 3.3).

As no automated CNV calls were made for IC AU101 and MK0038, a manual calling approach was used. For IC AU101, this indicated a deletion of exon 33, ~340bp in size, with the same 3' breakpoint coordinate of 6121116 (table 3.3). In the case of MK0038, no deletion was identifiable.

3.3.3.3 Detection of a novel 5' intergenic region (IR) deletion in two IC

During the validation analysis of IC with pseudogene region deletions (del Ex33-34), a novel upstream deletion was detected by the array in IC P9F3 II:1 and 1585PM. This 5' IR deletion had the same breakpoints in both patients, indicating a heterozygous deletion with a minimum size of ~4.6 kb and a maximum size of ~6.1 kb located around 11 kb upstream of the ATG initiator of *VWF* (see figures 3.7 and 3.8). These deletions are discussed in more detail in 3.3.8.

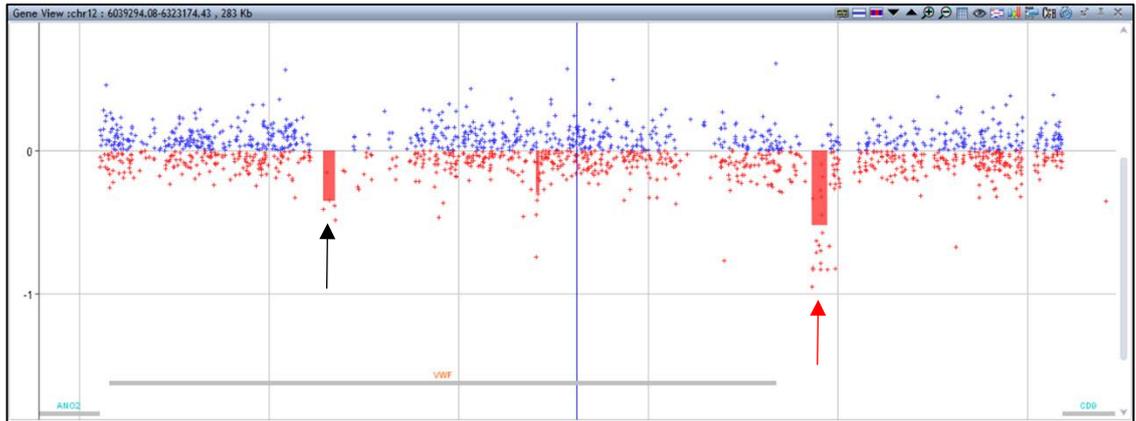


Figure 3.7 IC 1585PM pseudogene region deletion (exons 33-34). Black arrow indicates partial automated call of a deletion in the exon 33-34 region. Red arrow indicates the site of the 5' IR deletion.

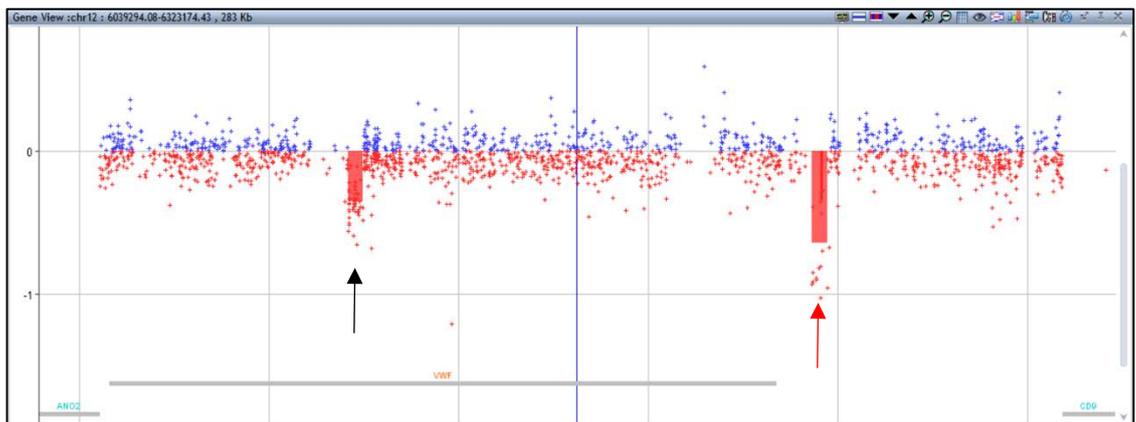


Figure 3.8 IC P9F3II1 pseudogene region deletion (exons 32-34). Black arrow indicates the automated call of a deletion of exons 32-33. Red arrow indicates the site of the 5' IR deletion, with the same breakpoints as IC 1585PM (Figure 3.7).

Patient ID	CNV	Size (bp)	Start breakpoint	End breakpoint	Analysis Method	Array CNV size(kb) (max)	Array start (max)	Array end (max)	Confirmation of deletion
1585PM	del Ex33_EX34	-	-	-	Default CGH Manual	8.7 (10)	6110998 (6110998)	6119733 (6121252)	Intron 34 only (del EX34)
AU0020	del Ex33_Ex34 + ?dup Ex14_Ex16	-	-	-		0.35 (0.9)	6121116 (6120861)	6121460 (6121788)	Ex33 only (EX33-pEX34)
EU P6F1 III2	del Ex33_Ex34	3432	6118343	6121775	Trial-1	0.35 (0.98)	6121116 (6120861)	6121460 (6121836)	Ex33 only (EX33-pEX34)
AU0101	possible del Ex33_Ex34 ?mosaic	-	-	-	Manual Manual	0.35 (0.6)	6121116 (6120861)	6121460 (6121460)	Ex33 only (EX33-pEX34)
MMK0038	small del Ex28 (indel)	33	6128378	6128411		None	NA	NA	No
EU P9 F3 III1	del Ex32_Ex34	5193	6119785	6124978		3.8 (4.1)	6121116 (6120861)	6124940 (6125002)	del Ex32_Ex33 (EX32-pEX34)

Table 3.5 IC with CNV in the pseudogene region. The array breakpoints and CNV sizes listed are those determined via the indicated automated detection protocol. Manual analysis of the maximum region taken from the two flanking probe coordinates are shown in italics in brackets. Partial deletions of exons are indicated by *p*.

3.3.4 Design iteration 2: 8x15k v2

The results obtained from the initial validation study using the 8x15k v1 array were used to direct a re-design of the array, aimed at improving CNV calling accuracy on problematic regions outlined in 3.3.3.1. This design also aimed to fill any other exonic and intronic probe gaps in the design. In summary, four separate probe groups were designed (5' intergenic region, exons 3-4, intron 5 and introns 30-34) using the genomic tiling method, with no exclusion of repetitive sequences or restriction sites. All probe groups were then exported and sequences were run through BLASTn. Any probes that returned hits outside of chromosome 12 (and chromosome 22 for pseudogene region probes) were deleted from the probe list. Probe sequences that returned no hits due to lack of specificity, such as highly repetitive sequences were also discarded. Remaining probes were then imported back into Agilent SureDesign as new probe groups and added to the existing array design, resulting in an additional 122 *VWF* specific probes from 1805 to a total of 1927.

4 of the previously analysed validation samples were reanalysed on the 8x15k v2.1 design; P9F11 I:1, P9F3 II:1, AU0020 and AU0101. A comparison of these results with the initial design is shown in table 3.4. Overall, the redesigned array on the 8x15k platform successfully improved the performance of the design in terms of improving accuracy and resolution of breakpoints called for known CNV. The increased probe density resulted in improved accuracy and coverage without obvious detriment to the previous design. For P9F11 I:1, the initial array design detected a deletion of ~2 kb within intron 3. Reanalysis on version 2 reduced the size of the deletion detected to 1.3 kb, but importantly this now included 142 bp of exon 3. A closer analysis of the array data suggested that a larger deletion would have been called were it not for the presence of a single probe (ID A_18_P19125360) within the 5' region of exon 3 that had a log₂ ratio of 0.299. This probe region spans the remaining 5' 23bp of exon 3, furthermore the next 5' flanking probe within intron 2 displayed a log₂ ratio of -0.213. Using this outermost deleted probe as the boundary would indicate a deletion of 1.5 kb that would include the entire 165bp of exon 3.

Using design version 1, the deletion of exons 32-34 in IC P9F3 II:1 was only detected as a deletion of exon 32 – partial deletion of exon 34 using the maximum

breakpoints. Analysis with version 2 increased the deletion size from 4.1 kb to 5.3 kb (max size) and involved an entire deletion of exons 32-34. For the two validation samples with deletions of exons 33-34 (AU0020 and AU0101), version 2 was less successful. While the maximum breakpoints identified using design version 1 indicated deletions of exon 33 and a partial deletion of exon 34, design version 2 called a smaller region of only ~ 250 bp located within intron 34. However it was noted that these arrays were part of a sub-optimal run where increased noise was noticeable.

Despite a number of improvements, noisy data sets were still produced and in some samples, unusual probe shift patterns were observed. For example, a false positive deletion of exon 47 was observed in a number of samples using this design (see figure 3.9 B). One possible explanation for these probe shifts is the possibility of a DNA sample quality issue. This is possible as DNA samples were obtained from various laboratories and have been extracted using various techniques. Furthermore, many of the DNA samples used in this study were extracted a number of years prior to this analysis. Numerous freeze thaw cycle over time may have had a detrimental effect on the DNA quality.

A second possibility for these aberrant probe shifts is sub-optimal normalisation/centralisation, which could disrupt the log₂ ratio calling of the aberration calling algorithm. A unique feature of this custom array is the high density of probes within a single genomic locus. The calling algorithms available in CytoGenomics are designed to analyse data from microarrays with an even probe spread across the genome. These designs have more genomic probes across the entire genome that can be used for normalisation purposes. It is likely therefore, that the high local probe density of this custom design may be compromising the efficacy of the normalisation algorithms. Therefore the option of moving to a larger format design, to enable the inclusion of more genomic backbone probes was explored.

3.3.5 Design iteration 3: 8x60k

In an attempt to improve normalisation and reduce probe noise in the array data, a third design iteration was developed on the larger format 8x60k capacity arrays. This increased probe capacity enabled the existing probe sets to be transplanted across and crucially, enabled the addition of genomic backbone probes that could be used for normalisation. These were selected from the pre-designed International Standards for Cytogenomic Arrays (ISCA) 60k backbone probe groups from across the genome. It was hypothesised that this design would allow more normal probes from across the genome to be used in the normalisation and consequently using fewer *VWF* specific probes, thus improving dye bias correction via the rank consistent probe selection algorithm.

Again, 4 validation samples were reanalysed on the 8x60k design: P9F3 II:1, AU0020, AU0101 and IA0080. A summary of these CNV results is shown in table 3.6. Overall, an improvement was observed in CNV calling and general normalisation also appeared improved. This improved normalisation was evident in the analyses of AU0101. As discussed in 3.3.3.2, no automatic detection of the exon 33-34 deletion was flagged up by the automated analysis. This was improved upon in design version 2 where a deletion in intron 34 was automatically called. However, in this design a false positive deletion in exon 47 was detected in this case and a number of other patients. Moving to the 8x60k design rectified the false call in exon 47 and the exon 33-34 deletion was automatically called as a deletion of exon33 – partial deletion of exon 34, representing a significant improvement over the initial design. A comparison of how each of the 3 array designs performed for AU101 is shown in figure 3.9. Similar improvements were observed in other samples (see table 3.5).

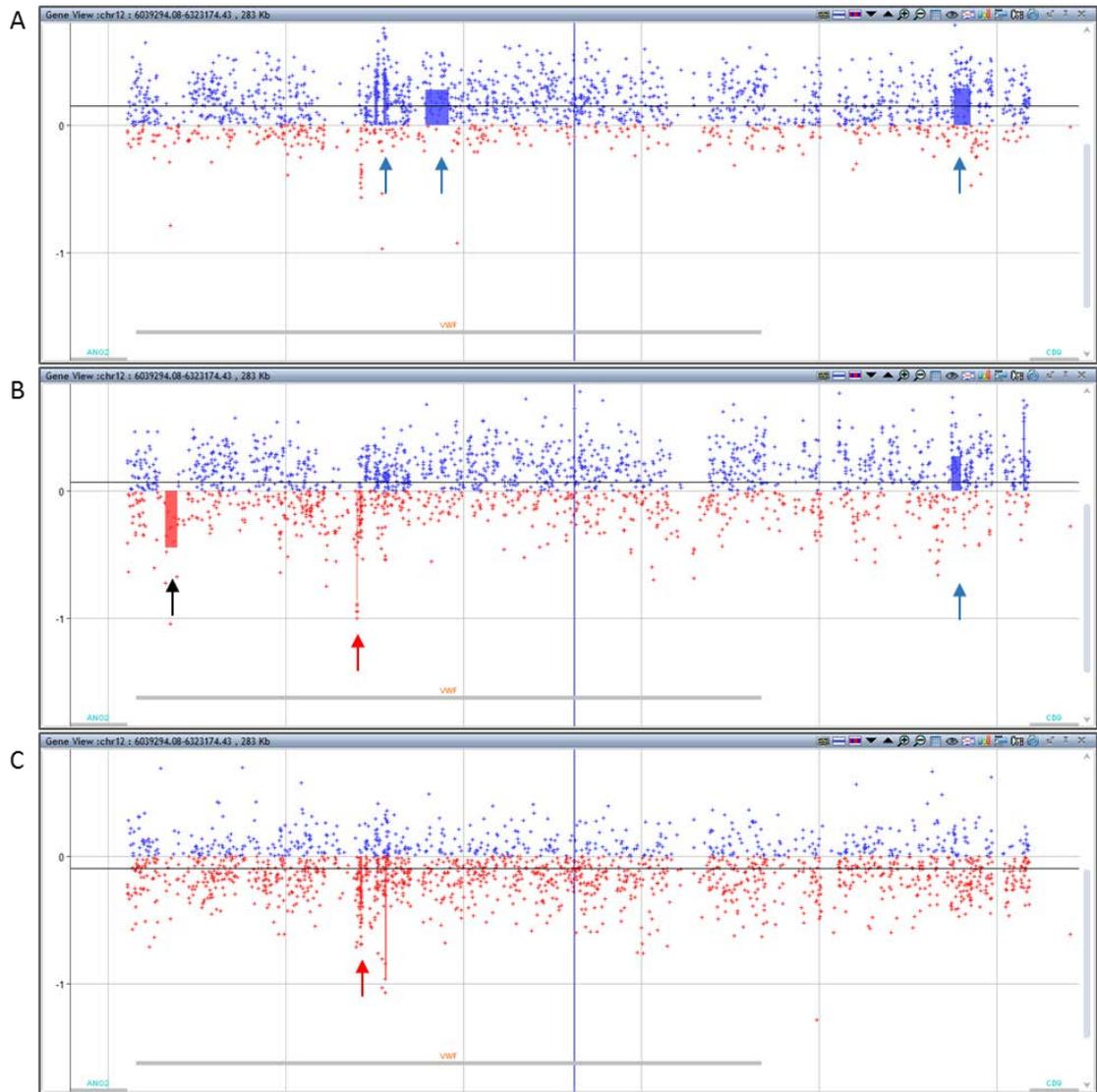


Figure 3.9. Example array profiles of the same IC, highlighting the differences between the 3 array designs. Here, results for IC AU101 are shown on the 8x15k v1 (A), 8x15k v2 (B) and 8x60k (C). The deletion of Ex33-34 is not automatically called in design v1, although probes in that region appear shifted (A). In design v2 del Ex33-34 is partially called by the software (red arrow in B). In the 8x60k design, a larger deletion is called including Ex33 and a partial del Ex34 (C, red arrow). Normalisation artefacts are observed in A and B (blue arrows). Design version 2 appeared to introduce a false positive deletion in exon 47 (black arrow). This artefact was removed in the 8x60k design.

Patient ID	Array version	CNV	Size (kb)	Start	End	CNV Confirmed
P9F11 I:1	-	del Ex3	2.1	6228754	6230840	
	8x15k V1		2	6228288	6230320	Intron 3
	8x15k v2.1		1.3	6229199	6230482	del pEx3
P9F3 II:1	-	del Ex32-34	5.2	6119785	6124978	
	8x15k v1		4.1	6120861	6125002	Del Ex32-pEX34
	8x15k v2.1		5.3	6119733	6125002	Del Ex32-34
	8x60k		5.5*	6119668	6125177	Del Ex32-34*
AU0020	-	del Ex33-34	-	-	-	
	8x15k v1		0.9	6120861	6121788	Del Ex33-pEX34
	8x15k v2.1		0.25	6119984	6120237	Intron 34
	8x60k		4.2	6117564	6121788	Del Ex33-34
AU0101	-	del Ex33-34	-	-	-	
	8x15k v1		0.6	6120817	6121460	<i>Del Ex33-pEX34 (manual call)</i>
	8x15k v2.1		0.26	6119984	6120241	Intron 34
	8x60k		0.83	6120954	6121788	Del Ex33-pEX34
IA0080	-	del Ex18	3.8	6150646	6154426	
	8x15k v1		3.3	6151137	6154452	Del Ex18
	8x60k		3.4	6151137	6154506	Del Ex18

** This call was split into two separate calls by the analysis software*

p = partial deletion of exon

3.3.6 Array based CNV screening in a type 1 VWD patient cohort

Following the initial validation experiments using design version 1, CNV analysis was performed on a type 1 patient cohort using the version 2 array initially and latterly version 3. 47 individual families, previously diagnosed with type 1 VWD, but for whom no causative mutation had been identified were analysed. From this cohort, exonic heterozygous CNV were detected in seven individuals. These comprised two exon 4-5 deletions, two exon 33-34 deletions and two deletions of exons 19-20. In one patient an apparent heterozygous duplication of exons 23-34 was detected.

3.3.6.1 Heterozygous deletion of exons 4-5

CNV analysis of patients IA00166 and MK0348 from the Milwaukee cohort identified a heterozygous deletion encompassing exons 4-5, with an approximate minimum size of 5.8 kb and a maximum size of 10 kb. Both cases had identical breakpoints (figure 3.10). Heterozygous deletions of exon 4-5 have been reported previously in type 1 VWD (Sutherland et al., 2009).

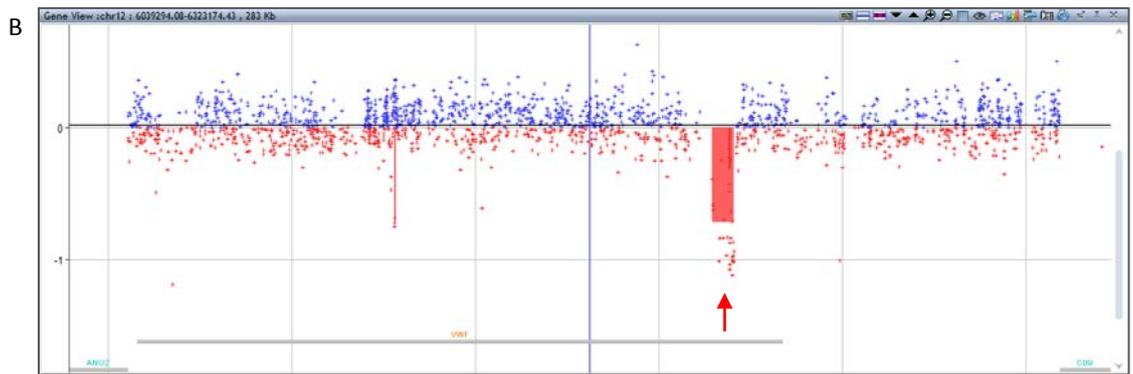
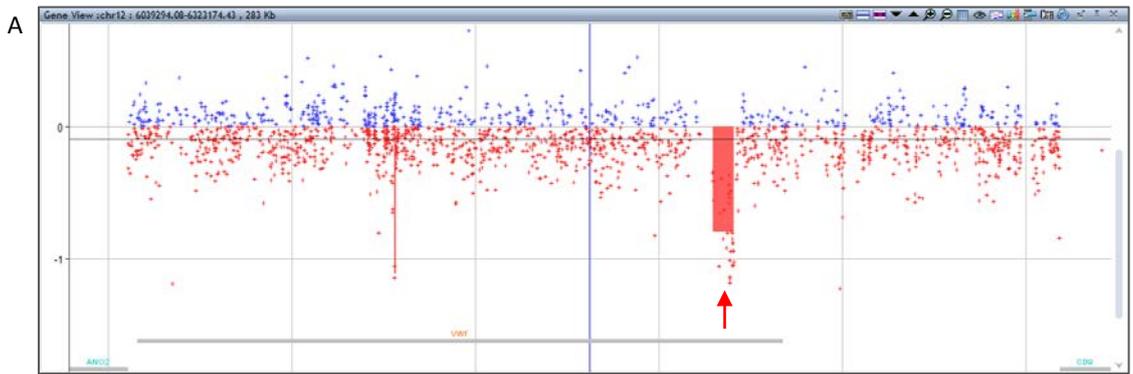


Figure 3.10 Detection of two heterozygous deletions of exon 4-5

Exon 4-5 deletions (red arrows) were detected in two patients from the Milwaukee cohort: MK0348 (A) and IA00166 (B).

3.3.6.2 Heterozygous deletion of exon 19-20

In two patients from the Canadian cohort, heterozygous deletions of exon 19-20 were identified (figure 3.11). In V037, the maximum size of the deletion called was ~4.7 Kb whereas in V372, the software called a larger region with a maximum size of ~4.9 kb. This is a relatively subtle difference and corresponds to two 3' flanking probes not being called in the smaller deletion whereas the 5' breakpoints were identical. In frame deletions of exons 19-20 have been previously reported in compound heterozygous type 3 patients (Boisseau et al., 2013, Bowman et al., 2013).

3.3.6.3 Heterozygous deletion of exon 33-34

Deletions of exons 33-34 were detected in two individuals; IA0048 and DT0143, both from Milwaukee and classified as type 1C (figure 3.12). This deletion has been reported previously in VWD patients (Yadegari et al., 2011, Boisseau et al., 2013, Cartwright et al., 2013, Theophilus et al., 2013). The deletions differed slightly in each case, with a maximum size of 4.7 Kb in IA0048 and 4.2 Kb in DT0143. This may reflect a slight difference in breakpoints at the 5' end as both 3' probe coordinates were the same. The discrepancy between the two samples at the 5' breakpoint region is accounted for by a single probe (ID A_18_P19124255) which is deleted in IA0048 but not in DT0143. The next 5' flanking probe used to estimate the maximum theoretical size of the deletion in IA0048 is ~450 bp further upstream. Whether this reflects a difference in the breakpoints of the two deletions is unclear and would need to be followed up by PCR and breakpoint sequencing.

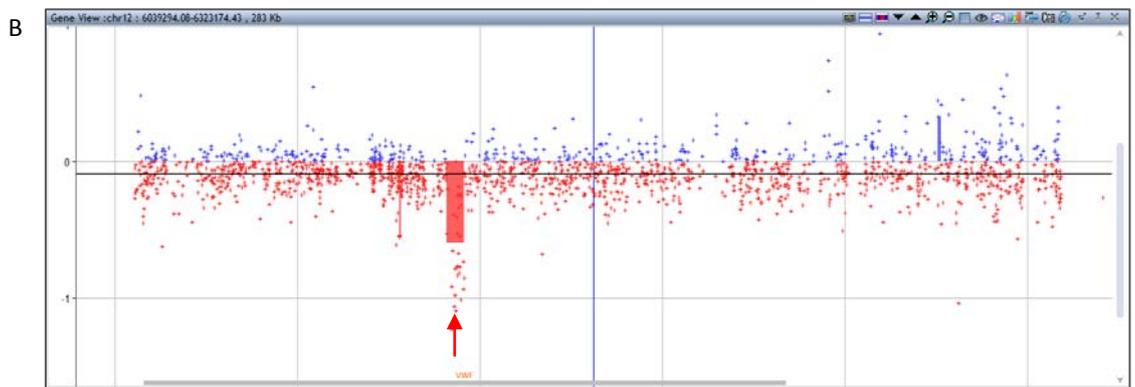
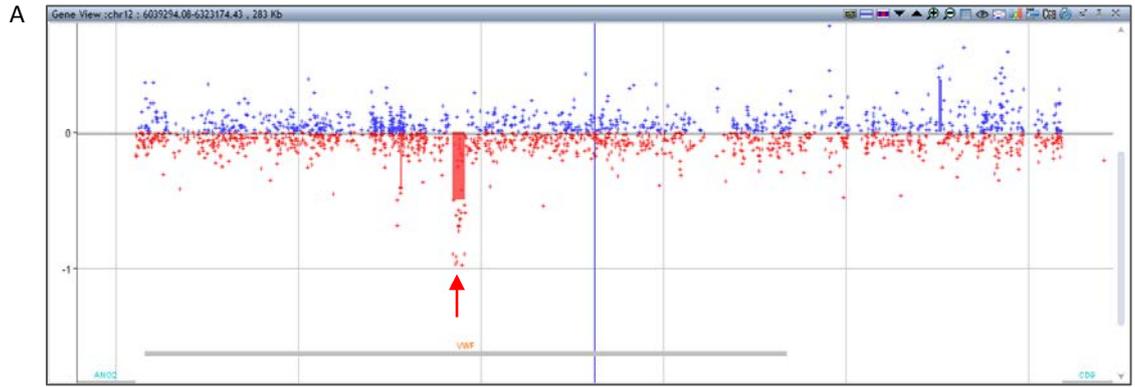


Figure 3.11 Detection of a heterozygous deletion of exons 19-20. Detected in patients V037 (A) and V372 (B). Both patient samples were obtained from Canada.

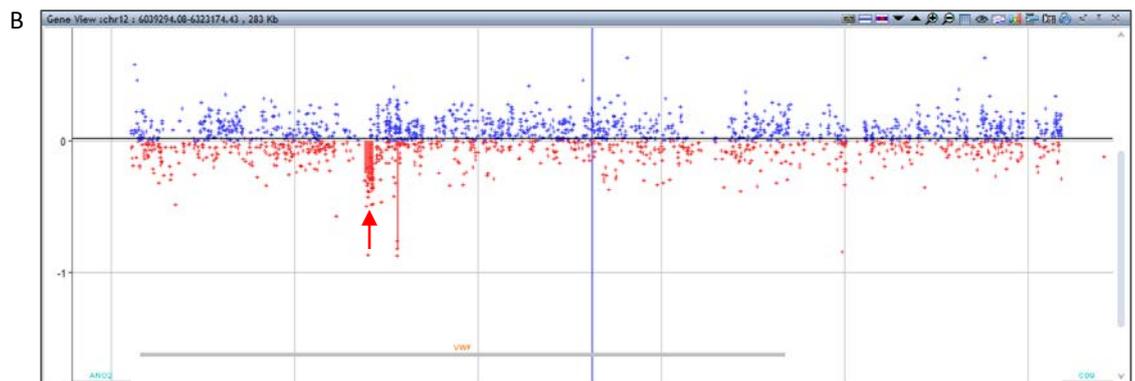
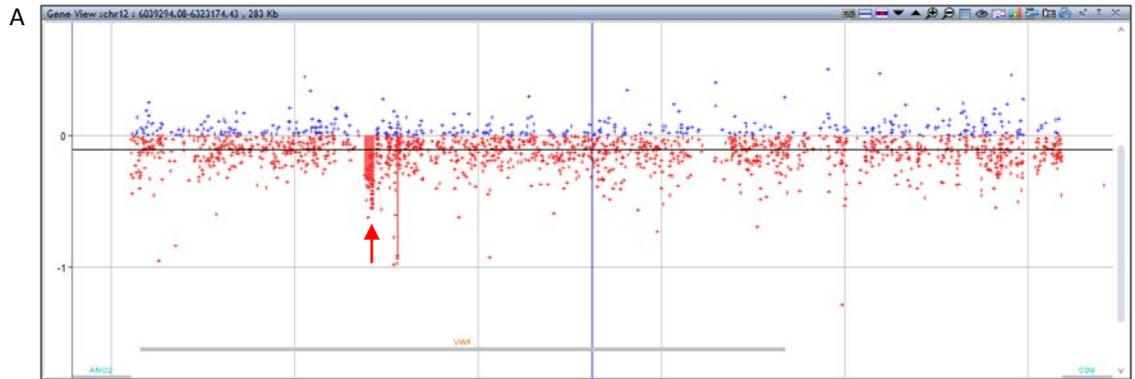


Figure 3.12 Detection of a heterozygous deletion of exons 33-34. Detected in patients IA0048 (A) and DT0143 (B). Both patient samples were obtained from Milwaukee.

3.3.6.4 Heterozygous duplication of exons 23-34

Analysis of NO153 identified an apparent 24 Kb duplication of exons 23-34, corresponding to the entire pseudogene region of *VWF* (figure 3.13).

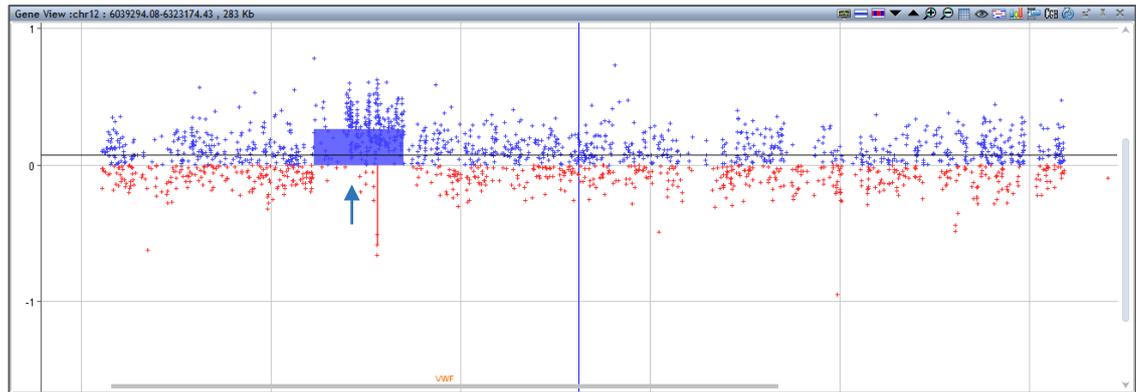


Figure 3.13 Detection of a pseudogene region duplication. An apparent duplication of exons 23-34 was detected in patient NO153 (Milwaukee).

3.3.7 Array based CNV screening in type 2 VWD patients

Only six type 2 patients were analysed for CNV in this study. Of these, only one individual was positive for CNV; patient MK0164. This patient was previously diagnosed with type 2A VWD and no candidate mutation had been identified. Analysis via aCGH detected a heterozygous 5' IR deletion further upstream than that described in 1.3.3.3. Both 5' intergenic deletions are discussed further in 3.3.9.

3.3.8 Array based CNV screening in type 3 VWD patients

A total of 14 VWD type 3 patients were analysed in this study. These comprised 8 cases submitted from the 3 WINTERS study, 4 cases from the Canadian type 3 study and 2 cases from Milwaukee. A summary of the findings is given in table 3.7 and the positive findings are discussed below.

3.3.8.1 3WINTERS type 3 cohort

A technical fault with the hybridisation oven, resulted in a significant drop in temperature during the processing of these samples. The result was numerous failed arrays and those that were acceptable, unfortunately suffered from excessive background noise. This is likely to have affected the identification of CNV in these samples. However, those presented here identified large CNV that were clear despite the noise in the data. In total, CNV were detected in six of the eight 3 WINTERS samples. In two patients from this cohort, whole gene deletions were detected. These included a 220 Kb heterozygous deletion in patient IT02-F02-II-P01 and a 220 Kb homozygous deletion in IT01-F19-I-P01. In patient DE02-F01-I-P05, a homozygous deletion of exon 6 had previously been identified. This was confirmed by aCGH and the result also appeared to show a heterozygous whole gene deletion, spanning ~ 244 Kb as shown in figure 3.14 C. However, this sample suffered severely from high background noise, suggesting that this apparent deletion was a false positive due to the poor quality data. Comparing the spread of probes with those in IT02-F02-II-P01 and IT01-F19-I-P01, particularly at the most 5' end of the region, i.e. those that were not called in the deletion demonstrates this point.

Patient ID	Allele 1	Allele2	CNV on aCGH	Start	End	Size (Kb)
IT01-F08-I-P01	Ex37 c.6551G>C <i>p.C2184S</i>	Unidentified	Het del Ex1-3	6227001	6261862	35
DE02-F01-I-P10	?	?	het del Ex1-5	6210061	6250699	40.6
DE02-F01-I-P05	Del Ex6	Del Ex6	Het whole gene del*	6059068	6303422	244
IT02-F02-II-P01	? del Ex1-52		Hom del Ex6	6199423	6218699	19.3
IT01-F19-I-P01	253 Kb del	253 Kb del	Het whole gene del	6055441	6276288	220
NL01-F01-II-P01	Ex45 c.7603C>T <i>p.R2535*</i>		Hom whole gene del	6055441	6275720	220
IT02-F09-II-P01	?	?	het 5' del-II	6255457	6260581	5.1
ES01-02-F02-I-P01	?	?	NA	-	-	-
T032	Ex25 c.3379 + 1 G>A	?	NA	-	-	-
T099	Ex15 c.1750_1765delinsCG <i>p.C584_S589delinsR</i>	?	None			
T108	c.4146G>T	?	None			
T143	Ex49 c.8043_8044delAG <i>p.R2681 fs</i>	?	None			
PH0012			Het del Ex19-20	6142263	6145762	3.5
CN0070			None			
			None			

Table 3.7. Type 3 VWD patients analysed in this study

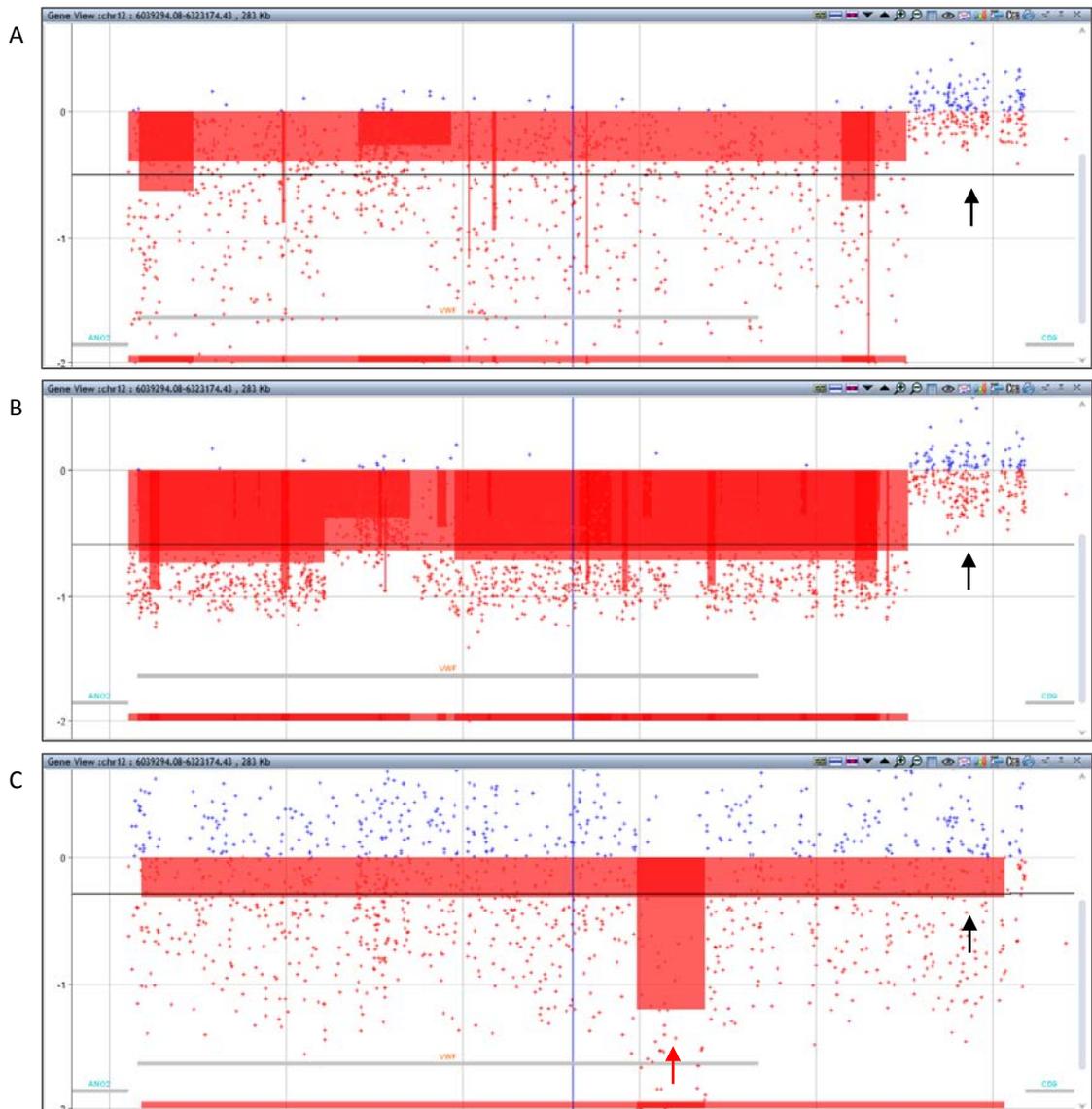


Figure 3.14 Detection of whole gene deletions in type 3 patients from the 3WINTERS study cohort. Homozygous 220 Kb deletion in patient IT01-F19-I-P01 (A), heterozygous deletion in IT02-F02-II-P01 (B) and a false call of a heterozygous deletion in DE02-F01-I-P05 (C). C also shows the real homozygous 19.3 Kb deletion of exon 6. Black arrows indicate the cluster of normal copy number probes, outside the deleted region in A and B. In C, the wide spread of probes is uniform across the entire region and unlike A and B, there are also a significant proportion of probes with a log2 ratio above zero. This is indicative of noisy, poor quality array data.

In the remaining three CNV positive 3WINTERS samples, smaller heterozygous deletions were identified, including multiple exons and an intergenic deletion. These included a deletion of exons 1-3 in patient IT01-F08-I-P01 (figure 3.15 A), who was previously shown to be heterozygous for a c.6551G>C (p.C2184S) mutation. A 40.6 Kb heterozygous deletion of exons 1-5 was detected in DE02-F01-I-P10 (figure 3.15 B) for whom no other candidate mutation had been detected. In the final patient, NL01-F01-II-P01, a heterozygous nonsense mutation in exon 45 (c.7603C>T; p.R2535*) had previously been identified but a second causative mutation could not be detected. Microarray analysis identified a heterozygous deletion of ~5.1 Kb upstream of VWF in the intergenic region (figure 3.15 C). Analysis of approximate breakpoints from the array data showed it to be the 5'del-II deletion that was detected in the type 1 cohort.

The proximal, 5' breakpoint regions of the deletions of exons 1-3 and exons 1-5 were different, with the exon 1-3 deletion extending a further 11 Kb upstream. Interestingly, this includes the 5'del-II deletion region although the approximate 5' breakpoints of these two deletions differ by ~1.2 Kb.

3.3.8.2 Canadian type 3 cohort

The four type 3 patients from Canada were previously analysed via sequencing and MLPA as part of the Canadian type 3 study (Bowman et al., 2013). Following aCGH analysis, a CNV was detected in one of these samples. IC T143 was found to be heterozygous for a deletion of exons 19-20, with a minimum size of 3.5 Kb and a maximum size of 4.7 Kb (array data not shown). These breakpoints were identical to those of the type 1 IC V037 (figure 3.11 A). This represents identification of the second causative mutation in this patient as they were previously found to be heterozygous for a frameshift mutation in exon 49 (c.8043_8044delAG; p.R2681fs).

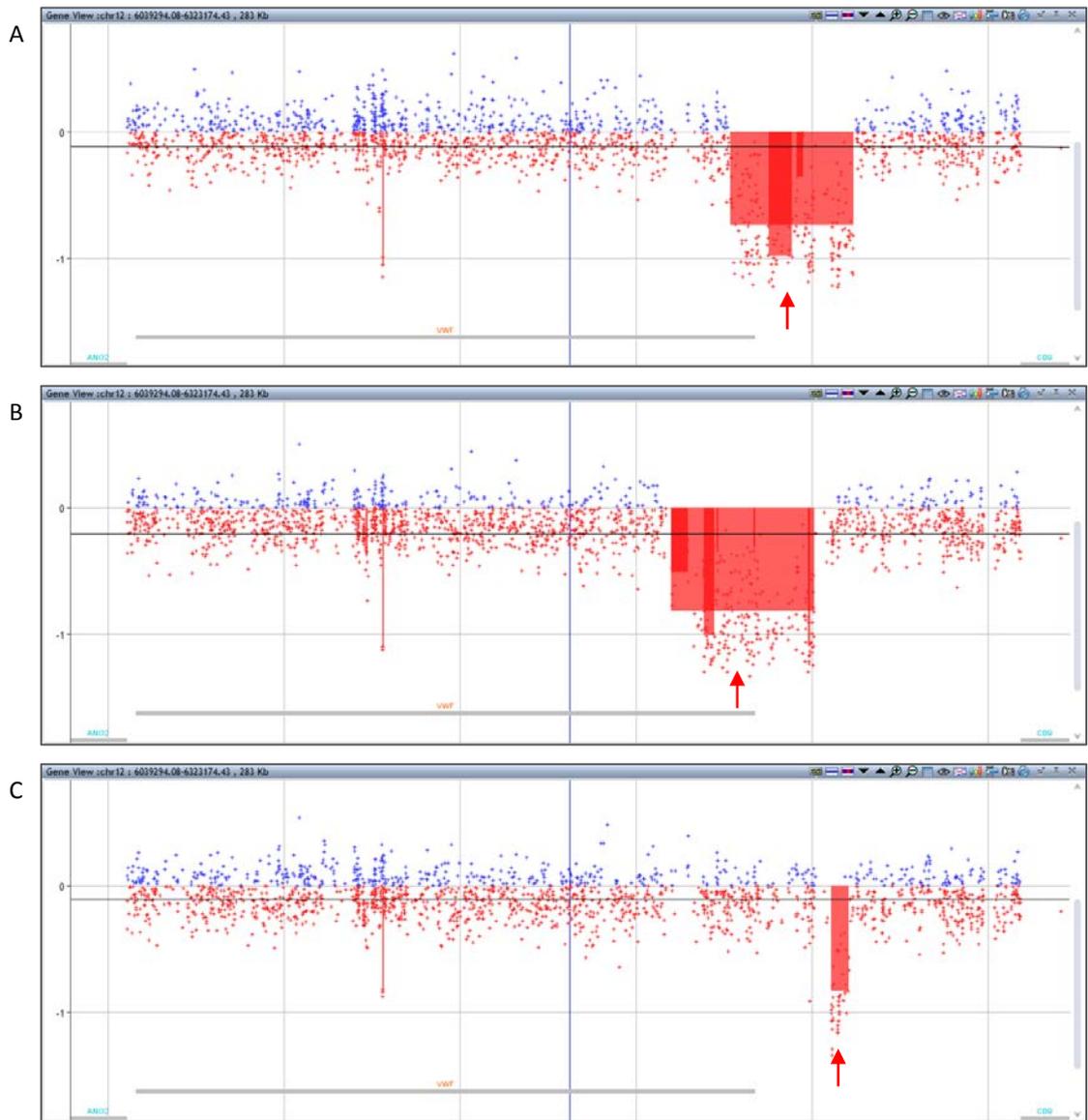


Figure 3.15 Detection of multi-exonic and intergenic heterozygous deletions in type 3 patients from the 3 WINTERS study cohort. Deletion of exons 1-3 that extended into the 5' intergenic region (A), deletion of exons 1-5 (B) and 5' del-II (C).

3.3.9 Detection and interpretation of two novel deletions in the *VWF* 5' intergenic region (IR)

As described in the validation study (3.3.3.3), array analysis of two IC in the pseudogene CNV cohort revealed the presence of a novel 5' IR deletion approximately 11kb upstream of the initiator ATG. This deletion was subsequently given the identifier 5'del-I and was ~4kb in size, with both patients showing the same breakpoint coordinates chr12:6243382 – 6247471 (Figures 3.7-3.8).

Following the validation study, the segregation of 5'del-I was analysed by performing aCGH analysis on a AFM and a UFM from P9F3 (figure 3.16). This demonstrated that while the 5' deletion segregated with VWD, it was on the same allele as the exon 32-34 deletion. Furthermore, AFM P9F3 I:3 was also found to have the 5'del-II in addition to 5'del-I. The UFM also inherited the 5'del-II allele suggesting that this intergenic CNV may not contribute to the VWD phenotype.

During the analysis of IC with no previously reported mutation, a second distinct 5' IR deletion was detected in AFM P2F12 III:1 and was identified as 5'del-II. This individual was initially selected for aCGH analysis because they had not inherited the p.Q1475* allele present in the IC and two other AFM of this family (see appendix 1). Upon identification of 5'del-II in this individual, aCGH analysis was performed on the two AFM and IC to determine the segregation of the 5'del-II CNV. This analysis demonstrated that 5'del-II did not fully segregate with the disease as it was not present in the other AFM and IC. This second 5' deletion was ~23 kb upstream of the ATG start site and around 5.1 kb in size.

As detailed above, these two novel 5' deletions were differentiated by labelling the first most 3' deletion 5'del-I and the second most 5' deletion 5'del-II. They are referred to as such from here on.

In the 47 type 1 families analysed, this study identified two 5'del-I (V396 and AU0122) and five 5'del-II. The two cases of the 5'del-I were the only identified CNV in those patients (see figure 3.17 A and B).

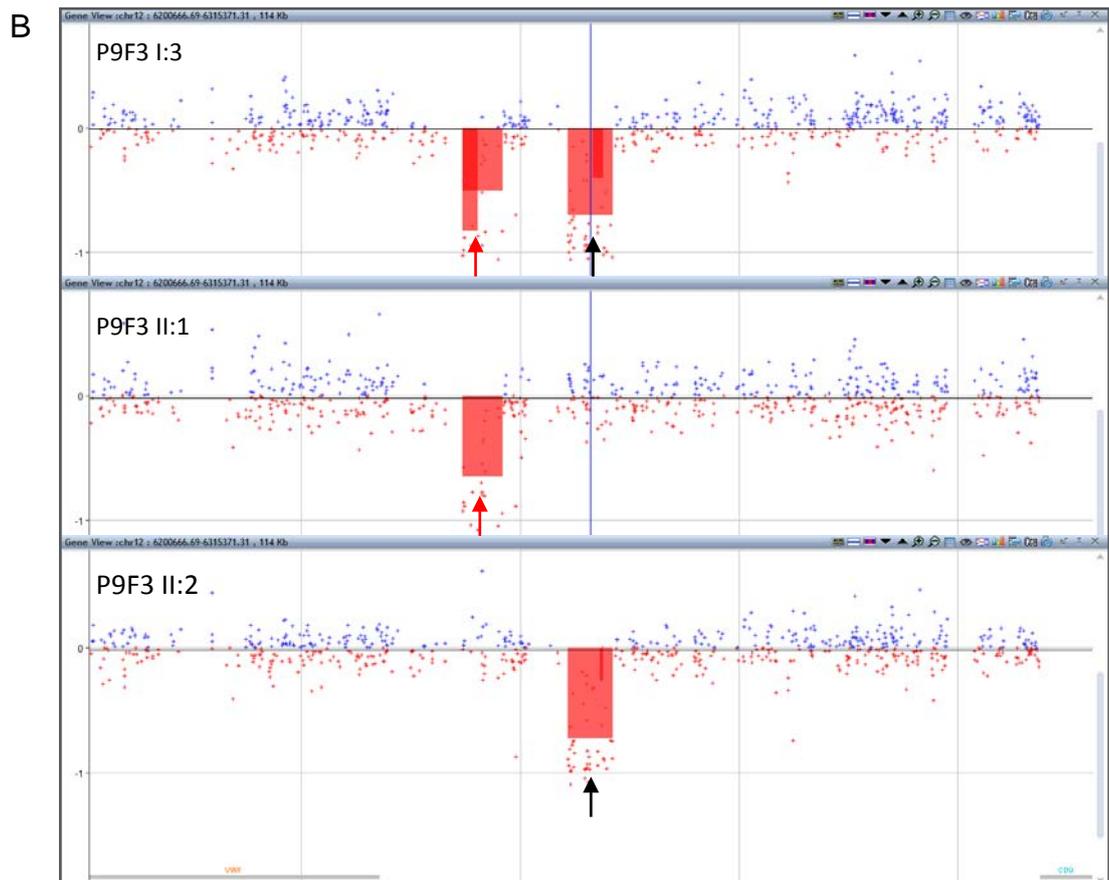
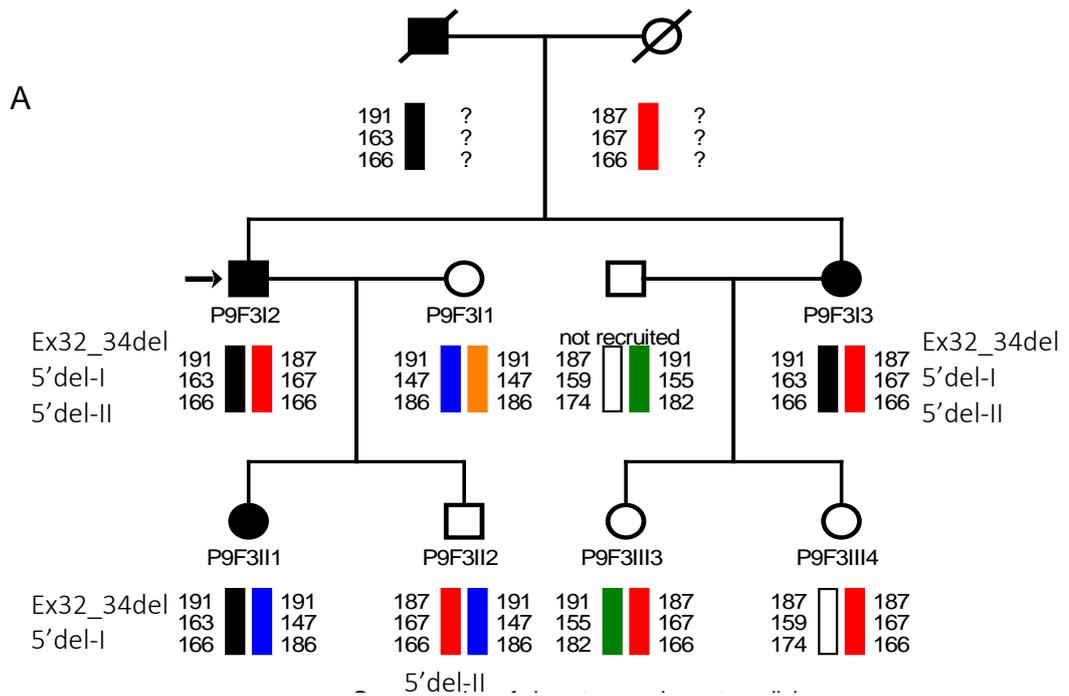


Figure 3.16. Detection of 5'del-I and 5'del-II in P9F3. A. Family pedigree for P9F3 showing segregation of del Ex32-34 and 5'del-I/II. **B.** aCGH detection of 5' deletions in this family. 5'del-I (red arrows), 5'del-II (black arrows).

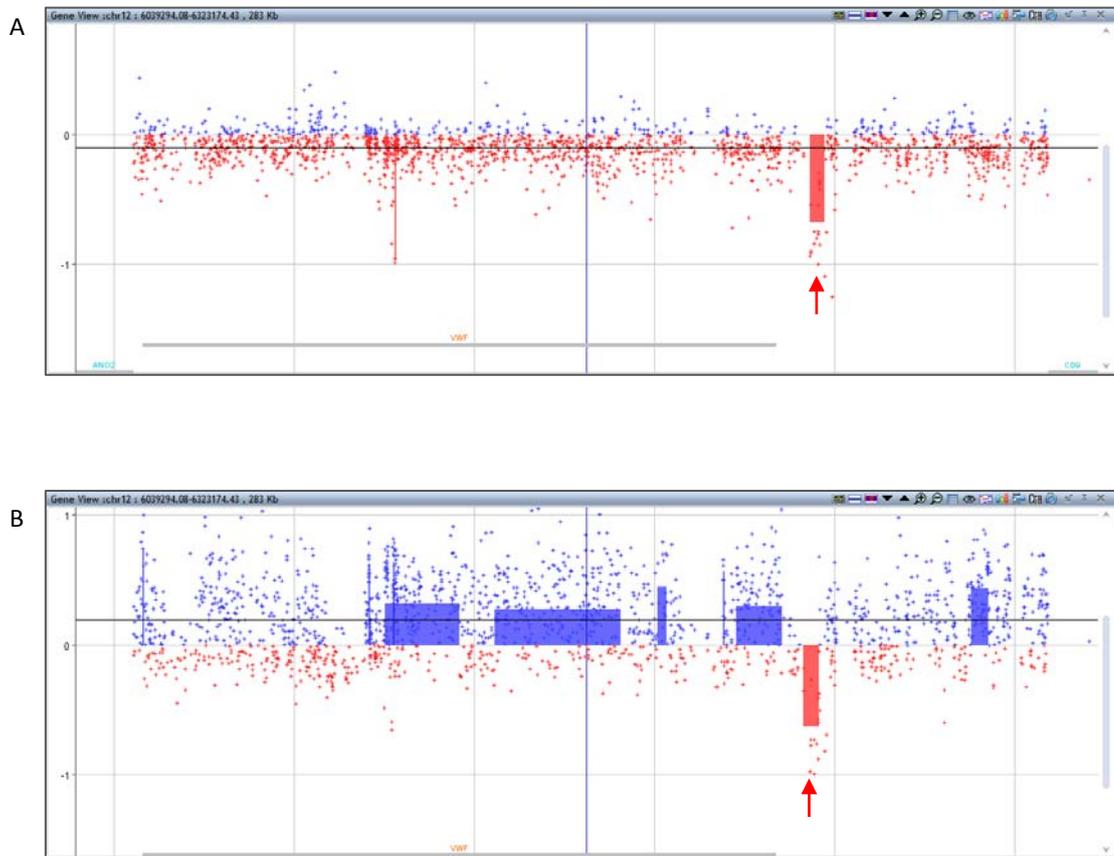


Figure 3.17. Two patients from the type 1 VWD cohort, where the only CNV identified was 5'del-I. Patient V396 sample was provided by Canada (A). Patient AU0122 (B) from the Milwaukee cohort was processed on a run where a technical issue with the hybridisation was reported. This sample had poor quality QC values resulting in numerous false calls due to the high background noise. Despite this, the 5'del-I deletion was clearly detectable.

In the two subsets of type 2 and type 3 patients analysed, one individual from each was found to be heterozygous for the 5'del-II deletion. While segregation analysis showed that 5'del-II was inherited by UFM, 5'del-I was only identified in type 1 IC and AFM and was not found in any UFM.

3.3.10 Incidental findings

As described earlier, one of the advantages of the aCGH approach was the inclusion of probes across the whole genome, enabling detection of potentially relevant CNV in other genetic loci. In two patients included in this study, incidental findings revealed deletions in other genetic loci. The first patient was mutation negative and diagnosed with type 1 VWD and also presented with a liver haemangioma. This patient was found to have a large 2.2 Mb heterozygous deletion on chromosome 2, deleting the entire coding sequence of the gene *FAM84A* (figure 3.18A). The second patient was also diagnosed as having type 1 VWD. In this patient an 83.6 Kb heterozygous deletion was detected on chromosome 3 (figure 3.18 B). This deletion was located in the coding region of the autophagy related gene *ATG7*.

3.4 Discussion

The aim of this chapter was to design a custom microarray that would enable high resolution CNV screening of the entire *VWF* locus, including exons, introns and 5'/3' intergenic regions. This investigation was driven by the hypothesis that in addition to identifying CNV not picked up by previous MLPA studies, aCGH would also enable the detection of non-coding CNV, which contributes to the VWD mutation spectrum.

The data presented here demonstrate the utility of aCGH in screening the entire *VWF* locus for CNV. Using this system, initial validation studies detected 9 of 9 known IC CNVs located outside the pseudogene region, with approximate breakpoints mapping to within a minimum of 26bp and a maximum of ~1.3kb of



Figure 3.18. Incidental CNV findings in two patients. A. Large 2.2 Mb deletion on chromosome 2 resulting in the deletion of *FAM84A*. B. An 83.6 Kb deletion on chromosome 3 within the coding region of *ATG7*.

previously mapped breakpoints. These are encouraging results and highlight a major advantage of aCGH in the identification of approximate breakpoints. The breakpoints can then be defined relatively easily by simple follow-up PCR, without the need for laborious and time consuming primer-walking strategies. The smallest mapped CNV available for analysis was ~2kb so it was difficult to assess the absolute resolution of this method. Out of 3 IC with known CNV but no breakpoint data, the array was able to confirm 2, indicating approximate breakpoints that can be followed up and confirmed by PCR. In the third, IC 2625SN, it is likely that the exon 22 deletion was too small to be detected. No information was available on the size of this deletion, however exon 22 is just 147bp in size and therefore the deleted region could be well below the array detection limit. This issue was probably compounded by the poor DNA quality of this sample as determined via the DNAOK! assay (Figure 3.1, sample B2). Interestingly, recent findings in the development of oligonucleotide microarrays have demonstrated that very small deletions in the range of 4 – 100 bp can be detected and this is dependent upon probe density (Belfield et al., 2014). This idea could be investigated in any future design modifications.

During the validation of the initial design some differences were observed when comparing known breakpoints with those detected by the array. For example the exon 3 deletion in IC EU P9F1111 was called as a 1.3kb intronic deletion on array version 1, but mapped breakpoints correspond to ~2kb and includes the entire exon. This result suggested insufficient probe coverage in this region and therefore this issue was addressed in array design v2, resulting in a smaller region being called by the analysis software but importantly detecting a partial deletion of exon 3. Furthermore, a manual analysis of the apparent maximum breakpoints in addition to disregarding a single aberrant probe resulted in a clearly identifiable heterozygous deletion of exon 3 with intronic breakpoints. Despite some modest improvements, design v2 appeared to suffer from a number of false positive calls in some samples. This could have been due to the inclusion of lower quality probes in the design that were included in an attempt to improve coverage. It would appear therefore, that there is a clear trade-off between maximum coverage and performance, particularly in a gene such as *VWF* consisting of numerous exons flanked by highly repetitive introns. In an

attempt to rectify this issue, a third design was generated by transferring the existing probe groups to the larger format 8x60k array design format. This enabled incorporation of more backbone probes from across the genome, thus providing higher proportion of high quality and importantly, normal copy number probes for the software to use in the normalisation process.

This design evolution resulted in a custom microarray that performed accurately and reliably even in the detection of CNV in the pseudogene region. The presence of the highly homologous, partial *VWF* pseudogene on chromosome 22 was anticipated to be problematic for aCGH analysis, as the array cannot differentiate between pseudogene and gene. While a deletion outside this region is recognised as a change from 2 copies to 1 copy (50% reduction), in the pseudogene region the array has to detect a change from 4 copies to 3 (25% reduction), making the ability to make these calls more sensitive to poor quality data. Despite this, promising results were still obtained for CNV detection in the exon 23-34 region in the initial design validation. Previous targeted array studies have suggested that identification of small CNVs can be possible with careful manual analysis of aCGH data (Askree et al., 2013). Initially, a similar approach was required to identify the subtle probe shifts that occur in the *VWF* pseudogene region. In IC 1585PM, manual analysis suggested a ~10kb deletion encompassing exon 34 but not exon 33 as previously identified. While it did appear that the deletion extended further upstream into intron 32, a cluster of 3 probes located within exon 33 did not show a negative log₂ ratio, thus either suggesting that the deletion does not encompass exon 33 and 34 as previously thought or that this cluster of probes are of poor quality. The breakpoints for this validation sample have not been mapped previously and therefore the result from this array validation may suggest that it is in fact an intron 33 to intron 34 deletion only. Unfortunately, this patient was not reanalysed on the v2 and v3 designs due to insufficient sample DNA. Future breakpoint mapping analysis may enable clarification of this finding.

For the remaining IC with exon 33-34 deletions used for validation, maximum breakpoints only correlated with a deletion of exon 33, all having the same 3' breakpoint coordinate of 6121116. The exon 32-34 deletion in IC EU P9F3I1, showed a deletion of exons 32 and 33, with the same 3' breakpoint. Careful

analysis of the data revealed a probe gap of ~305bp, 3' of the last shifted probe up to the next flanking probe, which included exon 34. The addition of extra probes in this region during the design modifications improved the detection of deletions involving exon 34. This likely reflects the improved probe density, introduced in design 2 and the improved data normalisation afforded by the additional probes in design 3. This was demonstrated by the identification of exon 33-34 deletions in two individuals from the type 1 test cohort, which were called as full deletions of both exons, with intronic breakpoints by the analysis software. The reported deletion in exon 28 of IC MK0038 is very small at just 33bp and combined with the noisy data and pseudogene location, it is likely beyond the detection limit of aCGH.

Following aCGH analysis of 47 type 1 families with no prior causative mutation identified, CNV were detected in 14 (30%), 7 of these were exonic (50%) and 7 were intergenic (50%). The exonic CNV comprised two exon 4-5 deletions, two exon 19-20 deletions, two exon 33-34 deletions and one apparently novel duplication of exons 23-34. Duplication CNV in *VWF* are relatively rare with only two previous studies reporting duplications of exon 6 and exons 9-10 respectively (Boisseau et al., 2013, Obser et al., 2016). A finding of a pseudogene duplication is therefore potentially very interesting, however this finding is still to be confirmed.

Only a small number of mutation negative type 2 VWD samples were obtained for array analysis and the 5'del-II was detected in just one of these. In 14 type 3 VWD patients, CNV was detected in 7 (50%) and comprised one each of a heterozygous and homozygous whole gene deletion, a single homozygous deletion of exon 6, and one case each of a heterozygous deletion of exons 1-3, 1-5, 19-20 and 5'del-II. Figure 3.19 summarises the aCGH findings resulting from this study.

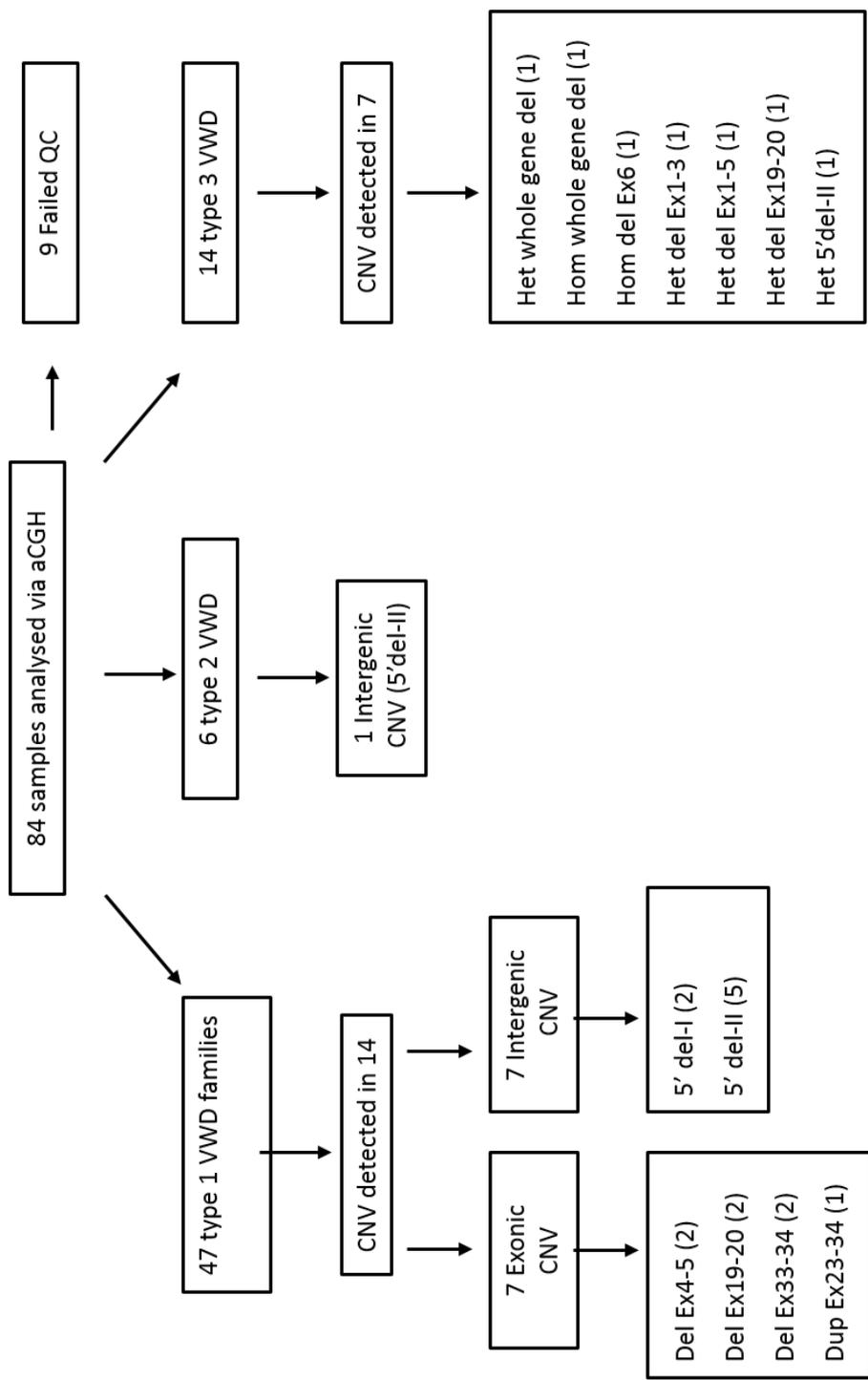


Figure 3.19. Overview of CNVs identified via aCGH in this study.

Although no intronic CNV were detected, intergenic CNVs were identified in nearly 15% of the type 1 VWD cohort. This intergenic CNV comprised two novel 5' intergenic region deletions, 5'del-I and 5'del-II, located ~11 Kb and ~23 Kb upstream of the *VWF* ATG initiator. Together, these two deletions accounted for 50% of the CNV identified in type 1 individuals.

In this study, 5'del-I was only detected in type 1 VWD IC and AFM. In one family it segregated with the exon 32-34 deletion, making inference to any pathogenic contribution difficult. However, in two mutation negative IC, 5'del-I was the only identifiable mutation. While it is possible that these patients may have other deep intronic mutations that as yet are unidentified, the results presented here do point towards a view that 5'del-I contributes to VWD in these patients.

The 5'del-II deletion was detected more frequently with 5 type 1 families and a single individual from each of the type 2 and type 3 subsets. The identification of 5'del-II in P2F12 III:1 is a potentially interesting finding as this patient does not have the p.Q1475X mutation that is shared by three affected family members. However, analysis of EU families P2F13 and P9F3 UFM showed the presence of 5'del-II in these individuals. The detection of this deletion in UFM and type 2 and type 3 individuals suggests that this CNV is unlikely to be contributing to the disease phenotype in VWD and may instead be a benign common copy number polymorphism.

As novel findings in a VWD cohort, however, both these 5' deletion CNVs warrant further investigation. Detailed analysis of breakpoints will give the true extent of each deletion and allow further *in silico* analysis of their functional genomic content and likely pathogenicity. These experiments and further analyses form the basis of chapter 4 of this thesis.

For those mutation negative individuals in which no CNV was identified, two possibilities exist as to the likely cause of their VWD phenotype. Firstly, prior sequencing analyses only analysed exons and exon/intron boundaries. Therefore it remains possible that single nucleotide variants and/or small insertion deletions exist in deeper intronic regions that have so far not been investigated. This hypothesis can be extended to include the existence of novel variants within the flanking intergenic regions. A second hypothesis is that these patients have either

CNV and/or mutations within or near other novel genetic loci that have some influence on VWF biology, particularly involving its biosynthesis, packaging and secretion. In this respect aCGH analysis can be a useful tool for rapidly screening the whole genome. This was exemplified by the detection of novel deletions in other genetic loci in two type 1 VWD patients. The first patient also diagnosed with a liver haemangioma was found to have a 2.2 Mb deletion on chromosome 2, resulting in the complete deletion of *FAM84A*. Little appears to be known about the function of this gene, however, it is reported to be upregulated in colorectal carcinoma and mouse liver tumours, increasing cell motility in both cases (Kobayashi et al., 2006, Kamino et al., 2011). In the second case, an 83.6 Kb deletion within *ATG7* was detected. This is potentially a very interesting finding as recent evidence has suggested a role for autophagy in regulated VWF secretion (Torisu et al., 2013). Indeed, in this study, the authors found that knocking down *Atg7* in mice resulted in fewer WPB, impaired VWF secretion and prolonged bleeding times.

The rapid progress in next generation sequencing (NGS) approaches should facilitate the identification of novel genetic loci type in the future. It should be highlighted that NGS approaches are also beginning to find use in CNV analysis with many diagnostic laboratories using in house algorithms and analysis pipelines to enable accurate detection of CNV (Hayes et al., 2013). Recent developments have demonstrated the utility of NGS data for the detection of single exon CNV (CoNVaDING) (Johansson et al., 2016). This may lead to an all-in-one analysis approach in the future where NGS provides a comprehensive analysis of single nucleotide variants, mutations and CNV across not only *VWF* but multiple genetic loci and eventually the whole genome. As the cost of NGS falls and analysis pipelines and workflows become more robust, this technique will likely become the method of choice for CNV screening across the whole genome.

In summary, this study describes the development, validation and utilisation of a custom *VWF* targeted microarray, able to detect changes in copy number across the entire *VWF* locus including introns, exons and 5' and 3' intergenic regions. In addition to identifying exonic CNV in a mutation negative VWD cohort, custom aCGH analysis was able to detect two novel 5' IR deletions that may have

important implications for our understanding of VWF levels in VWD and the normal population. These results highlight the potential utility of this array as a tool both in in VWD diagnostics and research, leading to improved understanding of the role of CNV in VWD pathogenesis.

4 Breakpoint mapping, *in silico* characterisation and screening of 5' intergenic deletions found at the *VWF* locus.

4.1 Introduction

In recent years there has been an intense effort to gain a better insight into the functions of the non-coding regions of the genome. This effort was accelerated with the initiation of the Encyclopaedia of DNA elements (ENCODE) consortium (consortium, 2012). This global collaboration was set up with the aim of constructing a comprehensive functional annotation of regulatory elements in the human genome. Following a pilot phase in 2003, these studies have generated a wealth of data from various epigenetic specific assays that are beginning to shed new light on our understanding of human genome structure and function. This is impacting many fields with the identification of novel aspects of gene regulation in health and disease. This chapter introduces aspects of the ENCODE functional annotation and its application to the understanding of *VWF* gene regulation. Specifically this chapter discusses the 5' intergenic regulatory landscape of *VWF*, the potential impact of non-coding CNV in this region and provides an analysis of the prevalence and effect of the two non-coding CNV deletions identified in chapter 3.

4.1.1 Non-coding DNA and ENCODE

The non-coding areas of the genome contain a variety of regulatory elements that are essential for the precise coordination of gene expression events. This tight control of transcription is achieved by various *cis*-acting elements including promoters, enhancers, repressors and insulators and also *trans*-acting factors such as transcription factors (TFs), long non-coding RNAs (lncRNA) and microRNAs (miRNA) (Maston et al., 2006). The ENCODE project has begun to compile a comprehensive list of these regulatory elements and factors, using a range of whole genome based assays. This has resulted in a valuable annotation of the non-coding genome, allowing identification of functional sites involved in gene expression regulation. Data from the ENCODE consortium studies suggest that up to 80% of the genome contains functional elements involved in some biochemical function (The 1000 Genomes Project Consortium, 2012). These sites are characterised by distinct genomic features within the regulatory landscape. Before discussing these genomic features or regulatory 'landmarks' it is necessary to briefly consider the structure of chromatin.

4.1.2 Chromatin, nucleosomes and histones

DNA within the nucleus of a eukaryotic cell is packaged in the form of chromatin. In the 1970's, chromatin was shown to consist of repeating units of two each of the four core histones (H3, H4, H2A and H2B) (Kornberg and Thonmas, 1974, Kornberg, 1974), around which the DNA is coiled. These 'beads on a string' were later coined 'nucleosomes' (Oudet et al., 1975) and represent the DNA-histone complexes connected by linker DNA. Prior to these structural studies, Allfrey *et al* had already demonstrated the post-translational modification of histones, specifically acetylation and methylation. This early work provided the first demonstration that histone acetylation was associated with increased RNA synthesis by RNA polymerase, thus linking histone acetylation to the removal of gene expression inhibition (Allfrey et al., 1964). Numerous histone modifications have now been described on various sites of histone proteins involved in the regulation of transcription, replication, repair and chromosome condensation at mitosis (Kouzarides, 2007). Histone modifications are able to govern gene expression patterns through establishing local and global chromatin environments. Of significance to transcription regulation are the acetylation (ac) and methylation (me) modifications of various lysine (K) residues.

Histone modifications may promote gene activation or repression. The trimethylation of lysines 27 and 9 on histone 3 (H3K27me3 and H3K9me3) are linked to gene repression (Barski et al., 2007) associated with the recruitment of the methyl-transferase Polycomb repressive complex 2 (PRC2) in a cell-type specific manner (Boyer et al., 2006). Conversely, mono-methylations of H3K27 and H3K9 are associated with gene activation possibly by blocking the spread of repressive marks such as H3K27me3. Other active chromatin-associated histone marks include H3K4me1 and H3K27ac. The mono-acetylation of H3K27ac is highly enriched at active enhancer sites, differentiating these regions from poised or inactive enhancers (Creyghton et al., 2010).

4.1.3 DNase-I hypersensitivity sites

A key feature of cis-regulatory elements is the accessibility of target DNA sequences by transcription factors or trans-activating factors. This accessibility is mediated by changes in the chromatin state, identifiable by an absence of nucleosomes, giving rise to open regions of chromatin with increased sensitivity

to nuclease cleavage. These DNase-I hypersensitivity sites (DHSs) were discovered over 35 years ago (Gross and Garrard, 1988) and have emerged as a valuable marker of regulatory elements in the human genome. Markers of chromatin state and cis-regulatory elements have played a key role in generating data for the ENCODE project. As advances in technology enabled high throughput screening and mapping of DHSs (Boyle et al., 2009), so came the first comprehensive DHS maps offering genome-wide DHS profiles across different cell and tissue types (Thurman et al., 2012). These are now a useful tool for functional analyses of non-coding regions.

4.1.4 Enhancers, insulators and silencers

Distant acting elements, able to increase basal levels of transcription in a cell type specific manner were originally identified in the SV40 viral genome and shown to enhance transcription of the β -globin gene, independent of distance and orientation (Banerji et al., 1981). This discovery led to the identification of the first mammalian enhancer, involved in lymphocyte-specific transcription of the immunoglobulin heavy chain gene (Banerji et al., 1983). Enhancers are now considered to be a fundamental component of eukaryotic gene expression regulation.

Data from chromatin-immunoprecipitation-based microarray experiments (ChIP-on-chip) as part of the ENCODE consortium have been used successfully to locate regulatory elements such as promoters, enhancers and insulators. Predicted enhancers in the human genome were found have characteristic histone modifications, being strongly enriched for H3K4me1 and frequently marked by H3K27ac, regions of DHSs and transcription factor binding sites (Heintzman et al., 2007, Heintzman et al., 2009). The importance of these cell-type-specific regulatory elements is also reflected in the level of sequence conservation at these sites. In comparison, promoter regions are marked by H3K4me3 at the transcription start site (TSS) and H3K4me1 downstream of the TSS.

Using ChIP-seq data generated by the Broad Institute, promoters and enhancers have been categorised as active/strong, weak/poised or inactive/Polycomb repressed, based on their chromatin structure and histone modifications (Ernst

and Kellis, 2010). This chromatin state segmentation track, available to view in the UCSC browser was generated by a Hidden Markov Model (ChromHMM) using data for nine factors (CTCF, H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K9ac and H3K20me1) across nine cell types in order to learn and model the combinatorial patterns of detected modifications (Ernst et al., 2011). These were then grouped into different chromatin states that are colour coded to reflect their predicted functional role.

Enhancer activity is controlled by the presence of boundary elements called insulators, which function to shield genes from non-specific regulatory activity such as that emanating from neighbouring genes. Insulator elements have been well characterised for the vertebrate β -globin gene locus. Chung and co-workers elegantly demonstrated that the 5' constitutive hypersensitive site of the chicken β -globin domain functioned as an insulator by its ability to insulate reporter gene expression (Chung et al., 1993). This activity was maintained in mammalian cell lines and in *Drosophila*. Furthermore, this region was shown to be functionally conserved in humans (Li and Stamatoyannopoulos, 1994). It is now understood that insulator elements achieve this regulatory partitioning through the recruitment of trans-acting factors such as the CCCTC-binding factor (CTCF) (Bell et al., 1999). More recent evidence suggests a role for CTCF in the three-dimensional organisation of chromatin interactions at the β -globin locus in a cell-type specific manner (Junier et al., 2012). Tissue specific CTCF boundary functions have also been shown for the human growth hormone gene locus (Tsai et al., 2014).

Gene silencers are targets for transcription factors that mediate repression of gene expression. These factors are known to play a key role in regulating developmental stage-specific gene expression patterns. Some of the clearest examples of these elements have been identified in the human β -globin gene cluster where silencing or repression of foetal haemoglobin expression is mediated by the binding of *BCL11A* (Sankaran et al., 2008)

Clusters or groups of these regulatory elements that control the expression of a specific gene or gene cluster are known as locus control regions (LCRs).

4.1.5 Non-coding CNV in disease

It is known that the majority of SNVs identified through genome wide association studies (GWAS), that are associated with increased risk of complex diseases, map to non-coding regions of the genome (Hindorff et al., 2009, Manolio et al., 2009). This finding suggests that it is likely that some of these functionally important variants are located within important regulatory regions of key genes. Similarly, the impact of CNV in non-coding regions has begun to emerge as an important factor in human disease. Much focus has been placed on developmental gene dysregulation, leading to congenital abnormalities (Klopocki and Mundlos, 2011), however, more recent studies have highlighted the importance of non-coding CNV in a wide range of diseases.

McCarroll *et al*, identified a common (10% frequency in HC) 20 Kb deletion upstream of *IRGM* associated with Crohn disease (CD). This work identified two distinct haplotypes in the population; the deletion, conferring CD risk and the reference (non-deleted), conferring a protective effect. Interestingly, each of these haplotypes exhibited distinct, cell type specific, expression patterns that had a direct effect on cellular autophagy, thus implicating the deletion allele as a causative variant in CD (McCarroll et al., 2008). Intergenic CNV may represent risk loci by causing dysregulation of a gene via disruption of its enhancer as reported for insulin growth factor binding protein 5 (*IGFBP5*), and breast cancer risk (Wyszynski et al., 2016). In some reports, intergenic CNV have shown to be a significant risk factor, despite the absence of any disruption to characterised regulatory elements (Julià et al., 2015). These findings may reflect the fact that regulatory element activity may only be detected in specific cell types under specific conditions, or that there is some positional effect of the CNV that alters the function of other regulatory elements.

The identification of small deletions within upstream non-coding regions have also enabled novel gene targets to be linked to certain diseases. A small 216 bp deletion upstream of the *KLC2* gene resulted in overexpression of *KLC2* in patients with spastic paraplegia, optic atrophy and neuropathy (SPOAN) syndrome. In this report, whole-exome sequencing failed to identify a causative mutation, while whole genome sequencing revealed the small homozygous deletion. Further work to knockdown and overexpress this gene in the zebrafish,

resulted in a mild to severe curly-tail phenotype, reflecting a similar neuromuscular pathology to SPOAN syndrome (Melo et al., 2015). This novel finding also demonstrated that a gain of function mutation resulted in a recessive condition. The fact that the deletion appeared to overlap the 5'UTR of *KLC2* and thus deleted an apparent promoter region suggested that this regulatory region may in fact function as a transcriptional silencer, whereby its deletion enables increased gene expression.

Dysregulation of gene expression may also be caused by gains in copy number of upstream regulatory elements. A heterozygous triplication of a 6 Kb region upstream of the matrix-metalloproteinase 19 gene (*MMP19*) was detected in all affected family members with cavitory optic disc anomaly (CODA), yet absent in 78 healthy controls and was not reported on DGV. The region was shown to have transcription enhancer activity in luciferase reported assays, thus linking *MMP19* overexpression in the optic nerve with CODA pathology (Hazlewood et al., 2015).

Non-coding CNV may also disrupt cis-regulatory silencers or repressor regions leading to disease due to aberrant expression of its target gene. This has been demonstrated by a deletion of an intergenic regulatory element that is essential for foetal haemoglobin silencing (Sankaran et al., 2011).

Transcription factors themselves are also susceptible to dysregulation via non-coding mutations and CNV. Mutations in a putative enhancer sequence 25 Kb downstream of the pancreas-specific transcription factor 1a gene (*PTF1A*) abolished enhancer activity, leading to isolated pancreatic agenesis. Interestingly one patient was found to harbour a 7.6 kb CNV deletion that included the entire enhancer region (Weedon et al., 2014).

4.1.6 VWF transcriptional regulation

Attempts to characterise the 5' flanking region of *VWF* following the publication of the gene sequence in the late 1980's, identified the *VWF* promoter region and transcription start site (TSS), TATA box element and a GT repeat motif (Collins et al., 1987, Bonthron and Orkin, 1988). Repetitive Alu elements were also identified upstream of the TATA box (Assouline et al., 1988).

These and subsequent studies have focused on a relatively small region of around 2 kb upstream of the *VWF* TSS. Functional analyses of the 5' flanking

region in *VWF* transcriptional regulation identified a minimal core promoter, two regions able to repress transcriptional activity and a 46 bp putative enhancer region located in the 5' region of an inverse Alu repeat element. These negative and positive regulatory elements were shown to interact in the regulation of *VWF* expression in a cell type-specific manner (Ferreira et al., 1993) and roles for SP1 and GATA TFs were proposed (Jahroudi and Lynch, 1994). This early evidence of the complex nature of *VWF* gene regulation led to further studies highlighting the roles of specific sets of TFs in *VWF* expression and repression. These included the identification of binding sites for the Ets family of TFs, present in the minimal core promoter sequence and required for promoter activity in endothelial cells (Schwachtgen et al., 1997) and examples of the transcriptional repression of *VWF* mediated through NF1 and the Oct family of TFs, binding upstream silencing elements (Jahroudi et al., 1996, Schwachtgen et al., 1998). From these studies, a complex picture of transcriptional regulation emerged, whereby repression of *VWF* is relieved in a cell-type and tissue specific manner. Via analogy to other studies, Schwachtgen *et al* hypothesised that putative cis-acting enhancer elements would exist either in intronic or 5' flanking regions further upstream or downstream from the core promoter region.

Attempts to identify additional cis-acting elements in *VWF*, have utilised a DHS directed approach (Kleinschmidt et al., 2008), reasoning that these open chromatin regions are markers of active regulatory elements. These studies identified a region in intron 51 containing cis-acting elements required for cell-specific expression of *VWF* in lung vascular endothelial cells. This mechanism has since been suggested to occur by chromatin looping, allowing the intron 51 element to interact with an NFI element ~100 Kb upstream and in close proximity to the core promoter sequence, resulting in de-repression of the promoter in lung endothelial cells (Nassiri et al., 2010).

In light of recent advances in our understanding of distant acting regulatory elements and the involvement of these regions in many human diseases (discussed in 4.1.5), it is reasonable to suggest that novel cis-acting regulatory elements exist, at greater distances from the *VWF* promoter, that play important roles in *VWF* gene regulation. It follows that novel CNV in these regions, such as

those described in chapter 3 of this thesis, may contribute to *VWF* transcriptional dysregulation.

Until now, no non-coding intronic or intergenic CNV has been identified at the *VWF* locus. Previous reports of mutations in regulatory regions are limited to the identification of a small 13 bp deletion in the basal promoter region, leading to disruption of Ets and GATA TF binding sites (Othman et al., 2010). The authors of this study noted that the potential contribution of a *VWF* LCR was yet to be explored.

4.2 Rationale and hypotheses

Previous studies have not analysed the 5' intergenic region of *VWF* for CNV. Therefore no previous reports of the prevalence or role of 5' intergenic CNV in VWD exist. The discovery of pathogenic non-coding CNVs in other diseases such as those discussed in 4.1.1 raises the question of whether non-coding variation contributes to the pathogenesis of VWD. In chapter 3 of this thesis, two deletion CNV (5'del-I and 5'del-II) were identified in a VWD cohort with a frequency of 2.3% for 5'del-I and 8.3% for 5'del-II. This corresponded to frequencies of 4.3% and 10.6% when taking the type 1 VWD cohort as a separate group. Furthermore, when incorporating the 18 known CNV type 1 validation samples, in which the 5' dels had not been previously identified, these frequencies in type 1 patients increased to ~6.2% for 5'del-I and ~12.3% for 5'del-II. Both these are significantly higher than those reported in phase 1 of the 1000 Genomes project (The 1000 Genomes Project Consortium, 2012), which reported these deletions as having frequencies of ~3.2% (5'del-I) and 1.8% (5'del-II). Therefore it was hypothesised that a lower frequency of these deletions would be observed in a control sample cohort, more in line with that of the 1000 Genomes. This would suggest that the deletion alleles are more prevalent in the VWD population and therefore are potential risk factors associated with reduced *VWF* levels. Although both deletions were reported in DGV using data from the 1000 Genomes Project, this data does not provide any phenotypic information on the individuals that carry these deletions. Furthermore, DGV and 1000 Genomes data do not specify the zygosity of each deletion.

4.2.1 Aims

Following the identification of two intergenic CNV in a VWD cohort and the rationale described above, the aims of this chapter were two-fold. Firstly, to map the breakpoints and perform an extensive *in silico* analysis of these two regions, in order to 1: Decipher their CNV deletion mechanism and 2: To ascertain whether these regions contained any genomic features characteristic of cis-regulatory elements. It was reasoned that if the deletions were likely to contribute to VWF level variation, then they would disrupt a regulatory element active in a relevant cell type such as HUVECs.

Secondly, because the aCGH results presented in chapter 3 indicated an increased frequency of these deletions in the VWD population, it was desirable to develop a means of high-throughput screening for these deletions, to determine their frequencies in a large control cohort.

Ultimately this would allow an assessment of the regulatory content of these regions, an understanding of whether these CNVs are more frequently found in VWD patients compared to controls and whether they correlate with lower VWF levels.

4.3 Results

4.3.1 5'del-I and del-II breakpoint mapping

A major advantage of aCGH is the accuracy of breakpoint approximations given by the two flanking probes as described in chapter 3. This enables a more rapid determination of breakpoints compared to techniques such as MLPA, which offer no positional information outside of the exonic probe regions. The result is laborious and time consuming primer-walking strategies using numerous primers to fill in the large gaps that exist between probes, to focus in on the true extent of the CNV. By using aCGH, particularly high probe density designs such as that described in chapter 3, breakpoints can often be determined from a single PCR using the flanking probe coordinates to design primers that correspond to the maximum theoretical CNV size.

Using this approach, two separate PCR assays were developed to enable amplification of the regions encompassing 5'del-I and 5'del-II. Initial experiments

focused on 5'del-I as this was the first of the deletions to be detected. This initial assay used primers designed to the minimum deletion size called by the analysis software, plus an additional 500bp of flanking sequence at both the 5' and 3' approximate breakpoints. Amplification of control samples (no 5' deletion) with OneTaq hot start polymerase resulted in no PCR product, suggesting that the deletion was larger than the array approximate minimum deleted region of ~4.6 Kb. A second assay using the maximum array approximate breakpoints (plus 500bp of flanking sequence) produced a faint band of ~7 Kb and a brighter smaller band of ~6 Kb. Further temperature gradient optimisations were unsuccessful at resolving this issue using the OneTaq polymerase as multiple non-specific bands were present across a range of annealing temperatures. It was reasoned that the expected 7 Kb product was too large for the OneTaq system to amplify, so the PCR was repeated on a temperature gradient using the long-range (LR) polymerase from KAPA Bio. The KAPA LR-PCR system uses a combination of *Taq* DNA polymerase and an engineered archaeal (B-family) DNA polymerase possessing proofreading capability. Initial results showed no specific amplification of the expected 7 Kb size across a range of annealing temperatures (48 – 65°C), with only non-specific bands appearing at lower temperatures and no bands being present over 53°C. Repeat experiments using a MgCl₂ gradient across the lower temperatures (48°C - 58°C) failed to reduce non-specific products and no correct size product was amplified. These results suggested a number of possibilities. Firstly, the region around the deletion breakpoints is highly repetitive and so potentially problematic for PCR amplification. Secondly, the polymerases used may not have been suitable for producing a relatively large PCR amplicon.

With this in mind, primers were redesigned using the maximum theoretical breakpoint coordinates with an additional 1000 bp of flanking sequence in the 5' and 3' directions. Using this method the expected amplicons in WT individuals for 5' del-I and for 5'del-II, were 6.9 Kb and 6.8 Kb respectively. PCR was then repeated using the Expand long template system from Roche using a gradient of annealing temperatures. For 5' del-I this resulted in a single band of ~7 Kb at 59°C, suggesting specific amplification of the 5'del-I genomic region. PCR was then performed on three individuals heterozygous for 5'del-I, resulting in a single

deletion specific product of ~300 bp (figure 4.1A). In these patients, no co-amplification of the WT allele was detected. The 300 bp products were purified via gel extraction and sequenced using the 5'del-I primers. Sequencing results were compared to the reference sequence and a clear deletion breakpoint was detectable with no insertion or complex rearrangements. These results indicated a deletion of 6.64 Kb, 9 Kb upstream of the *VWF* ATG initiator.

PCR analysis of heterozygous 5'del-II patients revealed a faint WT amplicon of ~6.8 Kb and clear deletion specific product in the region of 700 bp (figure 4.1B), indicating a deletion of ~6.1 Kb. As for 5' del-I, deletion products were purified via gel extraction and sent for sequencing to decipher specific breakpoint junctions. However, in this instance, a short stretch of readable sequence was followed by unreadable sequence. It was reasoned that the sequencing could be compromised by the long stretches of A and T repeats in the 5'del-II region, leading to a PCR deletion product consisting of a mixture of polymorphic repeat lengths. This could throw the sequence data out of sync leading to messy and unreadable sequence from which the breakpoints could not be determined. To overcome this issue, the PCR product of 5'del-II was cloned into the pGEM-T easy vector via TA cloning, transformed into NM554 competent cells and then sequencing of individual colonies was performed. Following this, the breakpoints could be clearly identified, however, due to the repetitive nature of DNA in this region it is possible that the exact breakpoint of this region may vary between individuals. This 6.1 Kb deletion was located 22.3 Kb upstream of the *VWF* ATG initiator. Breakpoint mapping of these deletions facilitated their *in silico* analysis and enabled two key questions to be addressed regarding these non-coding CNV. Firstly, to investigate the CNV deletion mechanism and secondly to perform a comprehensive functional analysis of the region in terms of its regulatory potential.

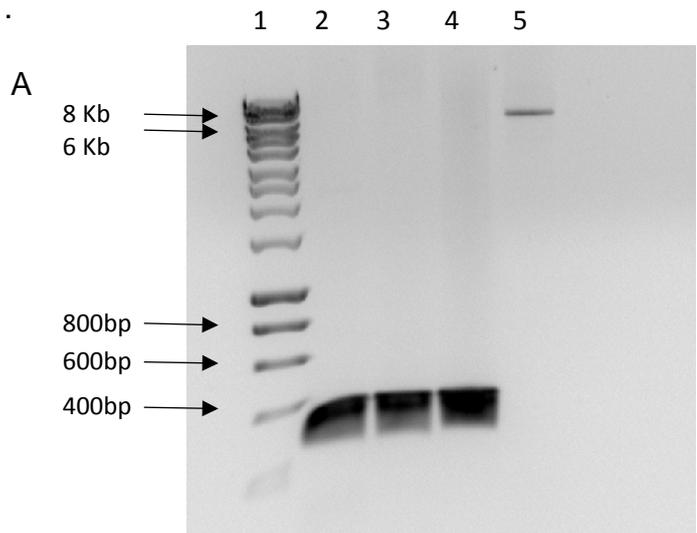


Figure 4.1 A. PCR across 5'del-I region Lane 1: Hyperladder 1 Kb size marker, lanes 2-4: Patients, heterozygous for 5'del-I resulted in a single deletion specific amplicon of ~300 bp. In lane 5, an individual with no 5' deletion (reference genotype) showed amplification of the full length region of ~7 Kb.

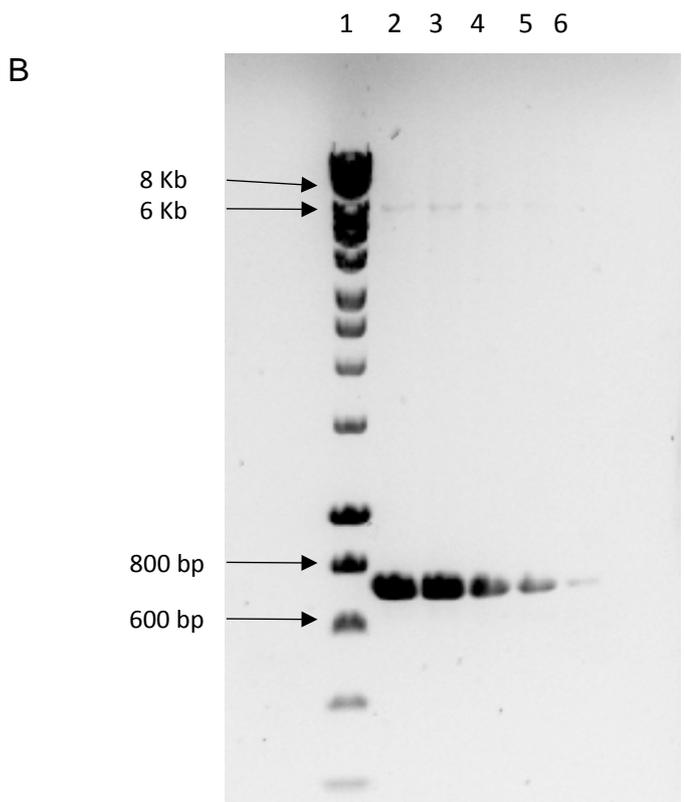


Figure 4.1 B. PCR across 5'del-II region Lane 1: Hyperladder 1 Kb size marker, lanes 2-6: Single patient heterozygous for 5'del-II, across a temperature gradient. Preferential amplification of the deletion allele is shown by the intense band at ~700 bp. This reaction was also able to amplify the reference allele, indicated by the faint band at ~7Kb.

4.3.2 *In silico* analysis of breakpoint architecture and determination of mechanism for 5'del-I

In silico analysis of repetitive elements was performed using the RepeatMasker track in UCSC (<http://genome-euro.ucsc.edu/index.html>). This identified the presence of a 160 bp AluSx family member SINE, that encompassed the 3' breakpoint of 5'del-I. However, analysis of the 5' breakpoint revealed no repetitive elements. Fork stalling due to DNA G-quadruplex secondary structure formation was discounted as a mechanism as no candidate G-quadruplex motifs were detected in the breakpoint flanking regions after analysis with QGRS Mapper (<http://bioinformatics.ramapo.edu/QGRS/analyze.php>, last accessed September 2016).

Further analysis of the breakpoint junction revealed the presence of a 2 bp AG microhomology region. This dinucleotide microhomology region was deleted at the 5' breakpoint and adjacent to the 3' breakpoint (figure 4.2). This suggested that the likely mechanism in this case was microhomology-mediated end joining (MMEJ). MMEJ always results in deletions and this is supported by the 1000 Genomes project data which detected no duplications in this region.

The DNA sequence flanking the breakpoints (75 bp in each direction) was also analysed for recombination-associated motifs using the DNA Pattern finder function in the Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/dna_pattern.html, accessed August 2016). The pattern finder utility identified two deletion hotspot consensus sites, DNA polymerase- α pause sites (x20 at the 3' region and x4 at the 5' region), a number of DNA polymerase frameshift hotspots (x2 3' and x7 5') and numerous topoisomerase sites (x30 3' and x37 at 5').

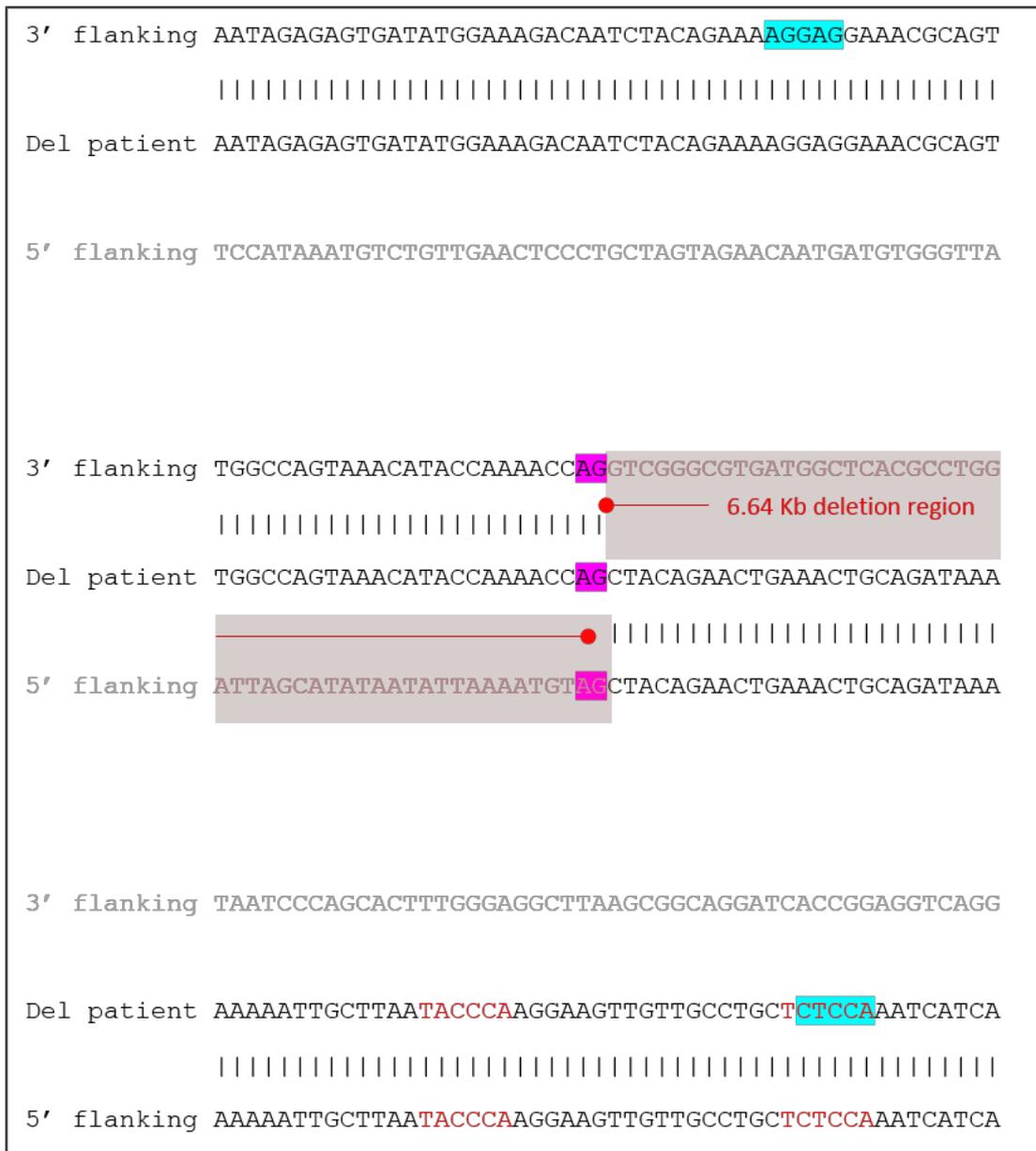


Figure 4.2 MMEJ deletion mechanism of 5'del-I. 75 bp of genomic sequence are shown either side of the 5' del-I breakpoint region. The AG dinucleotide region of microhomology is highlighted in pink. A region of AAA homology is also close to the breakpoint. DNA recombination hotspot motifs are highlighted in red text. DNA polymerase arrest sites are in blue.

4.3.3 *In silico* analysis of breakpoint architecture and determination of mechanism for 5'del-II

In silico analysis of the 5'del-II breakpoints revealed that AluY repetitive elements encompassed both breakpoints, suggesting Alu-mediated homologous recombination as the likely mechanism for this deletion. BLAST analysis of both AluY elements at the 5' and 3' breakpoints, revealed 90% homology and a run of 14 nucleotides at the breakpoint region that were identical in sequence (figure 4.3).

Analysis of recombination motifs around the breakpoint junction revealed numerous topoisomerase sites, a deletion hotspot consensus sequence, DNA polymerase α pause sites and DNA polymerase β frameshift sequences.

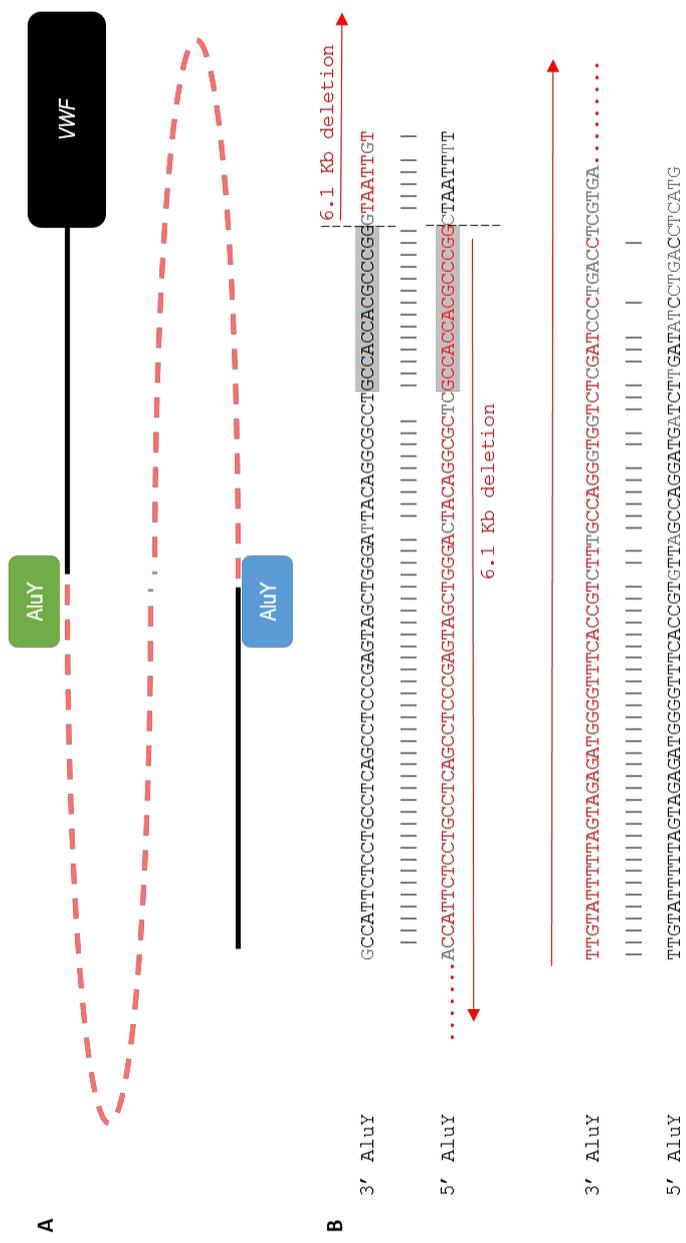


Figure 4.3 Alu-mediated NAHR mechanism in the formation of 5'del-II.

Schematic diagram of the Alu alignment between the highly homologous AluY elements leads to Alu-mediated homologous recombination (A). Alignment of the local sequence around the breakpoint junction (black dashed line) shows a region of 14 nucleotides with perfect homology (B). Deleted sequence is highlighted in red.

4.3.4 *In silico* analysis of the VWF 5' intergenic landscape

Prior to the analysis of the individual deleted regions, a survey of the entire VWF 5' intergenic flanking region was carried out (figure 4.4). Repeat masker analysis identified that of the ~80 Kb intergenic region, 52 Kb (~67%) is accounted for by interspersed repeat elements. This can be further broken down into 89 short interspersed nuclear elements (SINES), 33 long interspersed nuclear elements (LINES), 24 long tandem repeats (LTRs) and 13 other DNA elements. SINES are the most frequent class of repeats in this region, with Alu elements representing 25% of the 80 Kb intergenic sequence. To determine if this Alu enrichment had any functional or regulatory significance, the proportion of Alu elements that were within predicted enhancer regions was determined. In total, 45% of Alu elements overlapped a region classed as a weak or active enhancer.

Analysis of the enhancer content of the intergenic region revealed the presence of 34 strong enhancer elements. This cluster of strong enhancer regions is suggestive of a LCR.

4.3.5 *In silico* analysis of regulatory activity in the 5' del-I region

The ENCODE regulatory tracks data for enhancers, DNase-I Hypersensitivity sites and enhancer-associated histone marks were interrogated in the UCSC browser using the data available from HUVECs. Further analyses included assessment of vertebrate conservation and the presence of TF binding sites.

5' del-I was located ~9 Kb upstream of the VWF ATG initiator and was enriched for the enhancer mark H3K4me1, while the level of H3K27ac, although present in HUVECs was only at a relatively low level. The region displaying enrichment for H3K4me1 showed interspecies conservation and HUVEC specific DNase-I hypersensitivity sites, indicating that this was an active regulatory region.

The Broad Institute derived chromatin segmentation and classification track (chromHMM), was used to analyse the chromatin state within the deleted region. 5' del-I was found to encompass a 1 Kb strong enhancer element flanked by two smaller poised or weak enhancer regions. Interestingly, this strong enhancer region included a HUVEC specific DNase-I hypersensitivity site containing TF binding sites for c-Fos and c-Jun TFs. At the most 5' breakpoint of the deleted

region, a small 140 bp region of a second strong enhancer was also deleted. This region also included partial TF binding sites for CEBPB and c-MYC, however these were not associated with HUVECs and no HUVEC specific DHS was detected in this region. Other TF binding sites included NFYA, NFYB, MAFK and MAFF. Again these sites were not within HUVEC specific DHS regions and were located in regions with very low level or no active histone marks. Interestingly, NFY has been shown to inhibit VWF promoter activity in non-endothelial cells via by recruitment of histone deacetylases (HDACs) (Peng and Jahroudi, 2003). A detailed view of the region deleted in 5'del-I is shown in figure 4.5 and a list of the key regulatory features is shown in table 4.1.

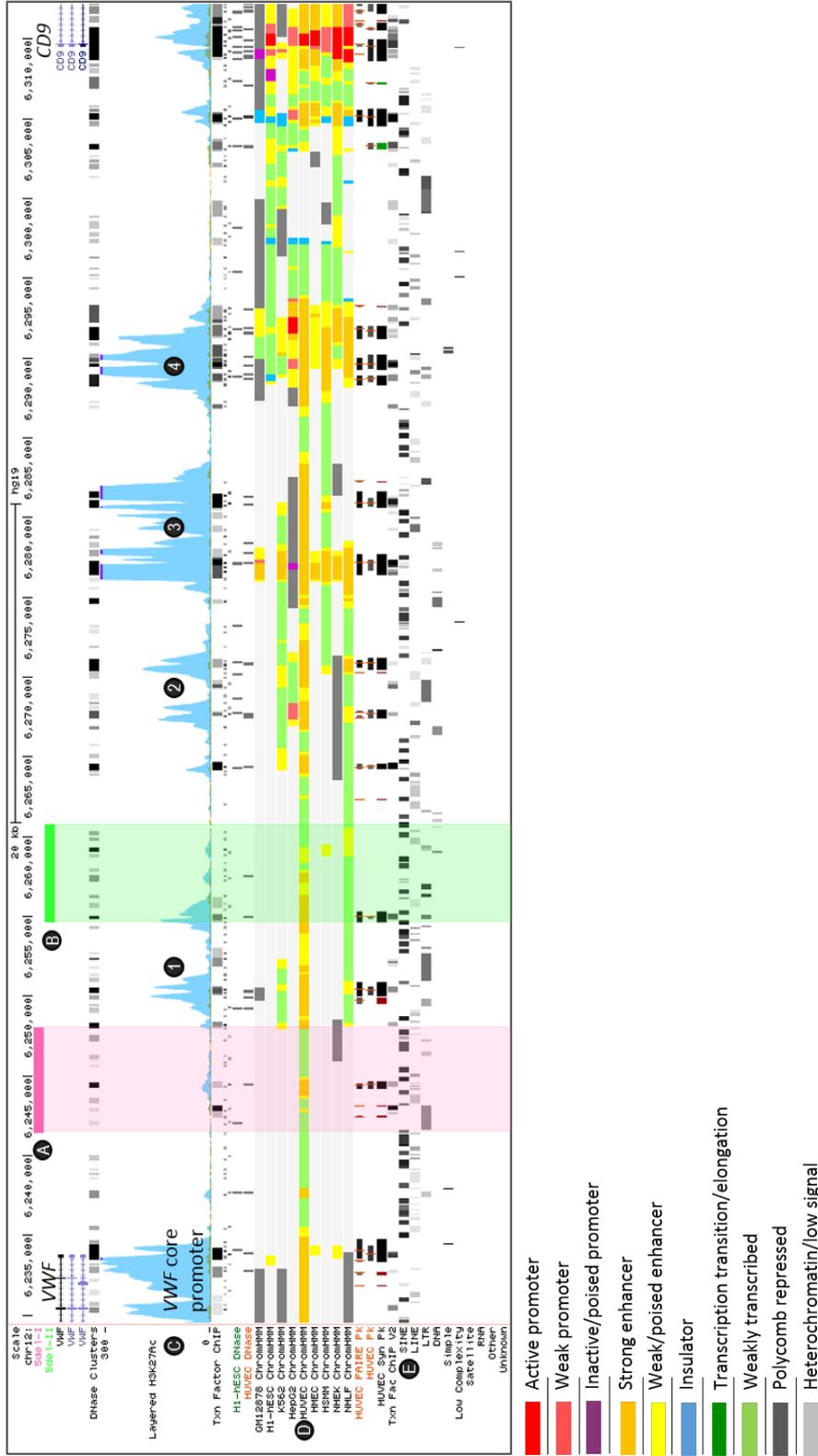


Figure 4.4.5' Intergenic regulatory landscape of VWF. The 5' intergenic region between VWF and CD9 is shown. The locations of 5' del-1 (pink bar vertical shaded region A) and 5' del-11 (green bar vertical shaded region B) are highlighted. The pale blue track labelled C, shows regions of H3K27Ac signal. In this lane, regions numbered 1-4 indicate other potential active regulatory regions upstream of VWF. Track D indicates the HUVEC specific chromatin states, colour coded as described in the key. Track E highlights the level of SINE repetitive elements in this intergenic region.

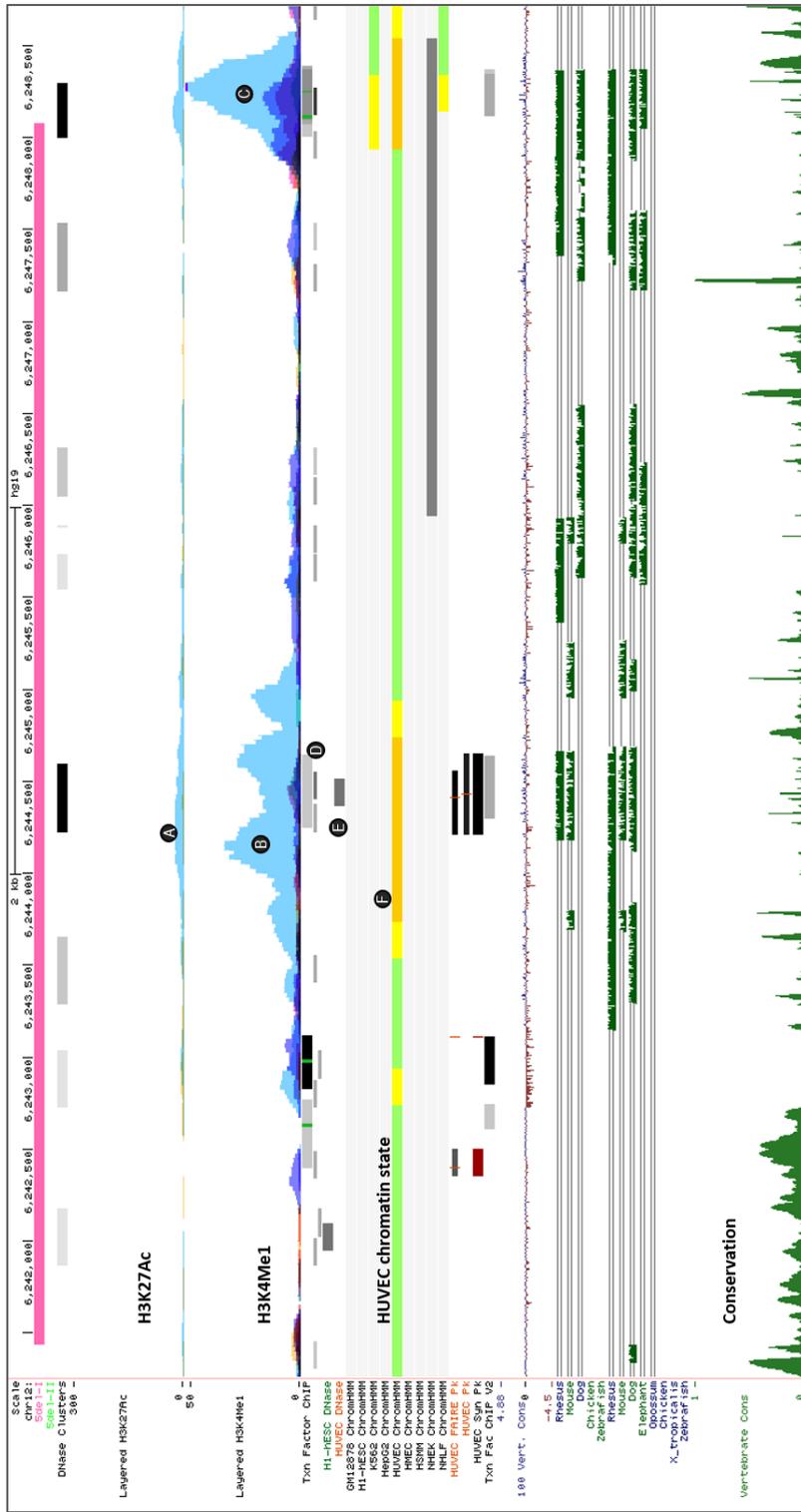


Figure 4.5 Detailed view of 5' del-I region. The extent of the 5' del-I deleted region is indicated by the pink bar (top). Low HUVEC H3K27Ac signal was detected in this region (A) while levels of H3K4Me1 mark were higher (B). A second region of H3K4Me1 signal was identified towards the 5' breakpoint (C). Regions of HUVEC specific DHSS and TF binding sites were present and in within a region of active enhancer chromatin state (D, E and F). Regions of conservation were present in this region but not within the regions of active regulatory potential.

Table 4.1 Markers of regulatory potential identified in 5'del-I

Regulatory marker	5'del-I
Dnase-I hypersensitivity site	8 clusters, 1 HUVEC specific.
Histone marks	Low H3K27ac, H3K4me1. Depletion of H3K4me3 and H3K9me3
TF binding sites	NFYA, NFYB, MAFK, MAFF, FOS, JUN , CEBPB, MYC
Conservation	Some, although not specifically at the active enhancer region

4.3.6 *In silico* analysis of regulatory activity in the 5'del-II region

Analysis of the 6.1 Kb region deleted in 5' del-II, revealed the presence of two regions of strong enhancer activity, associated with a relatively subtle increase in both H3K27ac and H3K4me1 histone marks. Both of these regions included a HUVEC-specific DHS region. The DHS site within the potential enhancer region with the highest level of H3K27ac signal included binding sites for the transcription factors GATA2, c-Fos, c-Jun and POLR2A. Similarly to the analysis of 5'del-I, these results supported the notion that the region deleted in 5'del-II contained functional regulatory elements of potential significance to VWF expression and levels. A detailed view of the 5'del-II region is shown in figure 4.6 and key regulatory features are shown in table 4.2.

Taken together and in the absence of any other intergenic CNV, the results provided by the *in silico* analysis of 5'del-I and 5'del-II provided sufficient evidence of regulatory function to warrant further investigation of these two intergenic CNVs.

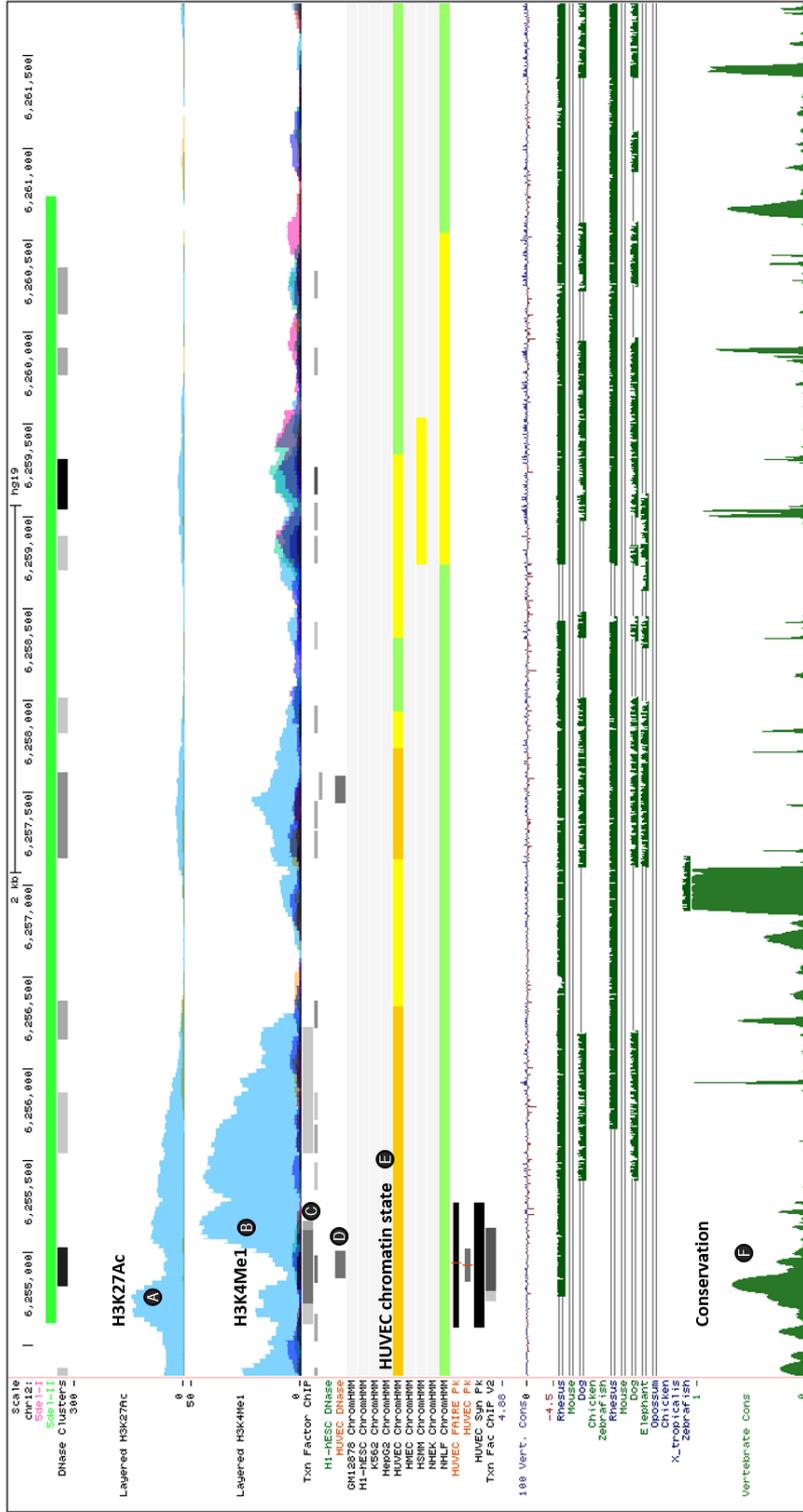


Figure 4.6 Detailed view of 5'del-II region. The extent of the deleted region is indicated by the green bar (top). At the 3' end of the deletion there are both HUVEC specific H3K27Ac and H3K4Me1 marks (**A** and **B**). Dips in these signals align with a HUVEC specific TF binding sites and DHS (**C** and **D**). The chromatin state in this region is classed as an active enhancer region (**E**). There is also significant vertebrate conservation at this region (**F**). Note the alignment of the DHS and TF binding site within the dip in histone mark signal strength. This reflects the depletion of nucleosomes at this site, making the chromatin accessible to TF binding, flanked by nucleosomes with acetylated or mono-methylated histones (peaks either side).

Table 4.2 Markers of regulatory potential identified in 5'del-II

Regulatory marker	5'del-II
Dnase-I hypersensitivity site	9 clusters, 2 HUVEC specific.
Histone marks	H3K27ac, H3K4me1. Depletion of H3K4me3 and H3K9me3
TF binding sites	GATA2, c-Fos, c-Jun, POLR2A (all overlap HUVEC specific DHS)
Conservation	Significant regions of vertebrate conservation, includes DHSS region

4.4 Development of a multiplex PCR assay for 5'del genotyping

Following the determination of deletion breakpoints for the two deletions and the *in silico* analyse of these regions described in 4.2, it was desirable to develop a multiplex assay that would enable a viable, rapid and cheaper alternative to aCGH to screen for both deletions. As described in 4.3.1, WT amplicons for the 5' deletion regions were produced in the analysis of individuals without the deletion, however, when the deleted allele was present in a heterozygous state, the smaller deletion product was preferentially amplified. The result was either no WT amplicon as in 5'del-I heterozygotes or a very faint WT product in 5'del-II heterozygotes. It was necessary, therefore, to design an additional forward primer for each deletion that annealed within the deleted region, thus producing a smaller, WT allele-specific amplicon when used in conjunction with the existing reverse primer for that deletion. An overview of the multiplex primer targets is shown in figure 4.7.

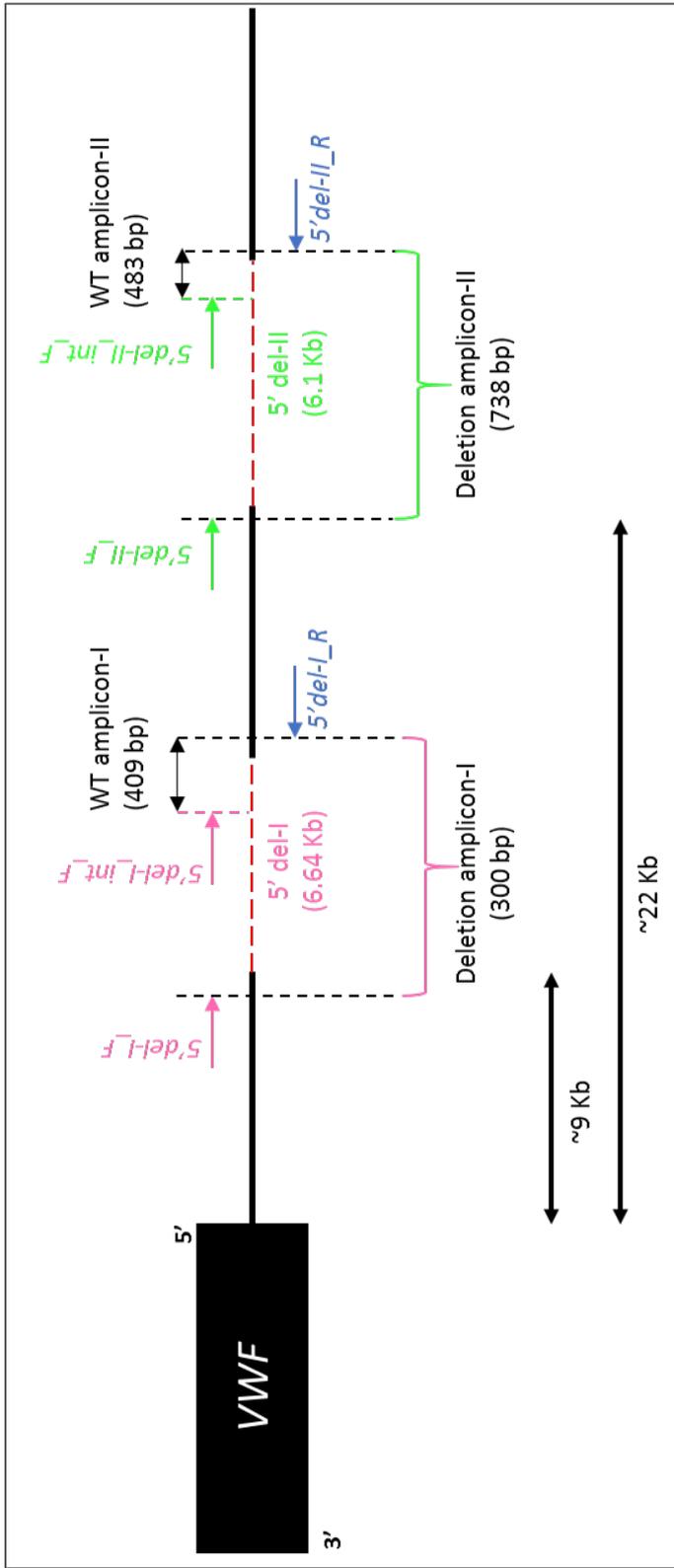


Figure 4.7 5'del multiplex PCR design. The multiplex contained 6 primers, 2 reverse primers and 4 forward primers (2 for each deletion). The deletion specific forward primers and reverse primers were designed to be located outside and flanking the deleted region. Therefore the resulting amplification would produce a small deletion band in the presence of a deletion (300 bp for 5'del-I and 738 bp for 5'del-II). The internal forward primers were designed within the deleted region, thus producing a short amplicon only in the absence of the corresponding deletion (409 bp for 5'del-I and 483 bp for 5'del-II).

4.4.1 Multiplex PCR optimisation

Initial experiments involved optimising the three-primer, dual amplicon PCR for each deletion separately, in order to assess the compatibility of the additional forward internal primers. These reactions were performed using the Reddymix 2x PCR reagent across a temperature gradient. Both of these single deletion multiplex reactions were successful, with 5'del-I producing a deletion specific product of ~300 bp and a WT product of 409 bp, whereas the 5'del-II multiplex produced a deletion product of 738 bp and a WT product of 483 bp. Optimum annealing temperatures appeared to be in the region of 58°C for 5'del-I and 56°C for 5'-del-II. However, in both cases the larger band was very faint with a significantly stronger band observed for the smaller product. Nevertheless, a 6-primer multiplex was tested with an annealing temperature of 56°C. This initial multiplex successfully amplified both WT products and the 5'del-I specific product. However, the larger 738 bp amplicon of 5'del-II was not produced, with non-specific products around 600 bp and 900 bp being produced instead. To address this issue two new alternative internal forward primers for 5'del-II were designed, however, these also proved unsuccessful. It was then hypothesised that the PCR reaction mixture was sub-optimal for the successful amplification of all six amplicons. In particular, because it was the largest product that had failed to amplify in the multiplex, it was reasoned that the smaller amplicons were amplifying at a faster rate, thus using up reaction components such as dNTPs etc. Rather than attempting to systematically optimise individual reaction components, less time consuming alternatives were sought. The multiplex was therefore repeated using the QIAGEN, multiplex master mix reagent. This resulted in successful amplification of all four PCR products, with only some faint non-specific bands appearing at lower annealing temperatures. As the reaction became cleaner as the annealing temperature was increased up to 62°C, a second reaction was performed with a gradient from 60°C - 66°C, with the highest temperature producing the most specific set of products. The number of cycles was also reduced from 35 to 30 to further reduce non-specific amplification. This optimised reaction was then taken forward and used to test four individuals that span the range of 5'del CNV combinations that may be detected (figure 4.8).

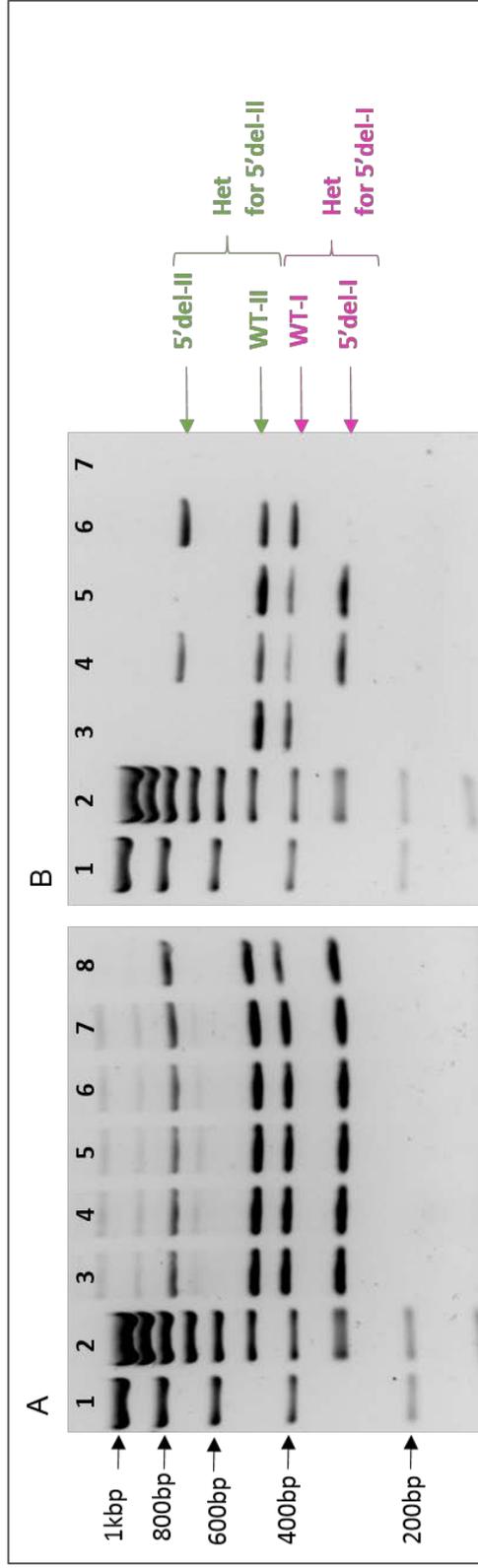


Figure 4.8 Optimisation of 5'del multiplex PCR and analysis of different 5'del genotypes.

A. The 6-primer multiplex was successfully optimised using the QIAGEN multiplex master mix. The reaction was optimised to an annealing temperature of 66°C (lane 8). **B.** Analysis of patients with 5'dels: no del (lane 3), both 5'dels (lane 4), 5'del-I (lane 5) and 5'del-II (lane 6).

4.4.2 5' del genotyping assays

Following multiplex optimisation, the forward primers were changed to FAM (5'del-I) or VIC (5'del-II) labelled versions, allowing the deletion and wildtype products for each deletion to be detected via fluorophore emission on the Applied Biosystems 3730 DNA Analyser. This enabled a rapid high-throughput approach for screening for the presence of 5'del-I and 5'del-II in numerous samples. To confirm the specificity of this reaction and to confirm that the addition of the fluorescent tag had not disrupted the PCR in any way, three patients were analysed that together spanned the range of genotypes that might have been expected: heterozygous 5'del-I, heterozygous 5'del-II and heterozygous for both deletions. An example of the fluorescent genotyping data is shown in figure 4.9.

4.4.3 Prevalence of 5' intergenic deletions in the healthy control cohort

The optimised 5' deletion multiplex genotyping assay was used to compare the allele frequency of the two 5' deletions in the type 1 VWD cohort and a larger HC cohort. From a European cohort, 958 healthy controls were genotyped for both 5' intergenic deletions. Analysis of this data revealed a significantly higher frequency than that reported by the 1000 Genomes project (table 4.3). 5' del-I was detected in 7.5% of the control samples whereas 5'del-II was present in 10% of samples. This finding was comparable to the frequencies observed in the small VWD cohort analysed via aCGH, suggesting that these deletions are not more or less frequent in the VWD population analysed in this study. Unexpectedly, a small number of individuals homozygous for each 5' deletion were detected in the HC samples (5'del-I Hom: n=3 and 5'del-II Hom: n=2). No obviously significant pattern or association of VWF levels was identifiable in these individuals.

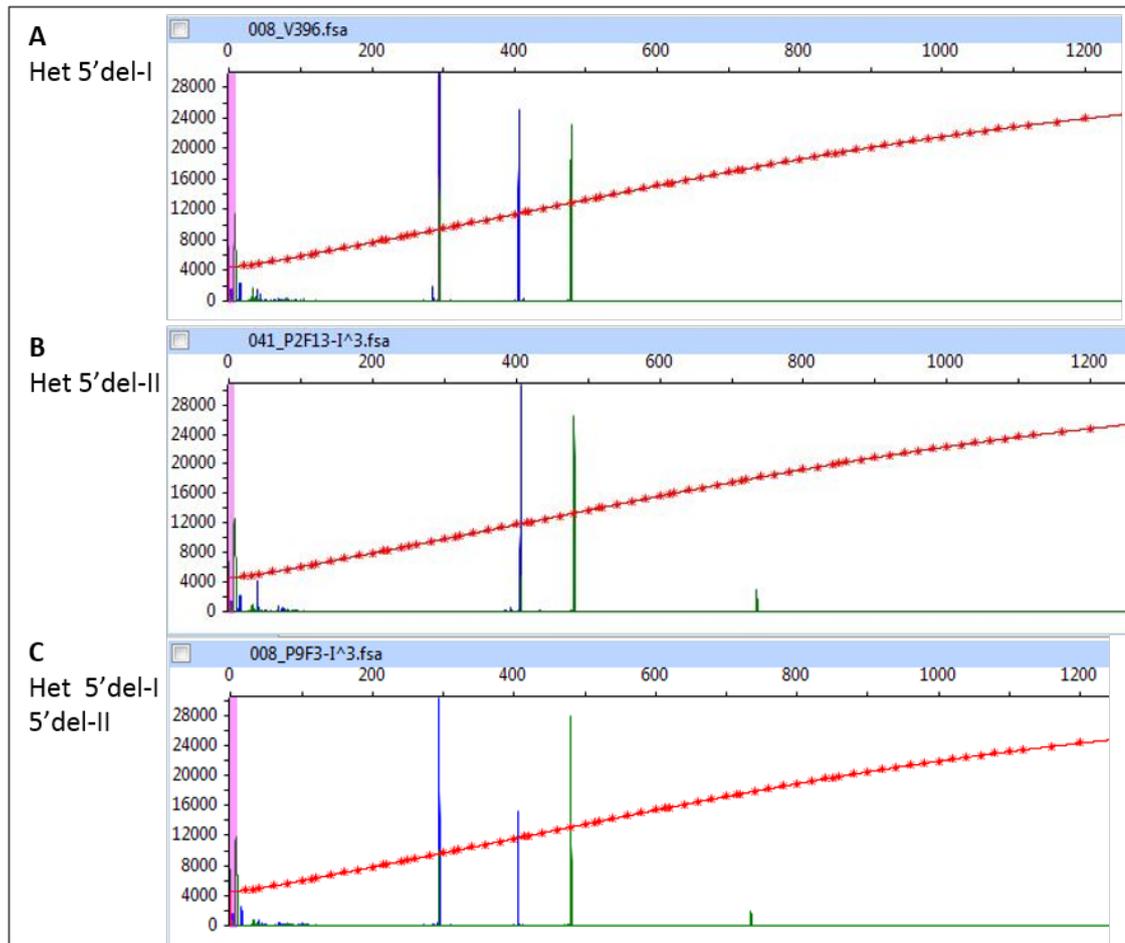


Figure 4.9 Fluorescent genotyping data for 5' deletions.

Forward primers for 5'del-I were labelled with a FAM tag, resulting in a blue WT peak at ~400 bp and a deletion specific peak at ~300 bp (A). The 5'del-II forward primers were labelled with a VIC tag, resulting in a green WT peak at ~490 bp and a deletion specific peak at ~730 bp (B). A patient with both the 5'del deletions is presented in C, indicated by the presence of all four peaks.

Table 4.3 Frequency of VWF 5' deletions in this study and 1000 Genomes

Genotype	Healthy Controls this study (%) (n = 949)	Type 1 VWD cohort this study (%) (n = 47)	1000 Genomes Phase 1 (%) (n = 1151)	1000 Genomes Phase 3 (%) (N = 2504)
5'del-I het	7.5	6.2	3.2	3.4
5'del-II het	10	12.3	1.8	3.8
5'del-I and del-II Het	0.4	0	NA	NA
5'del-I hom	0.3	0	NA	NA
5'del-II hom	0.2	0	NA	NA

4.4.4 Correlation of 5' deletions with VWF:Ag levels

Based on the apparent loss of active regulatory elements, it was expected that although the two deletions were present at similar frequencies in patients and HCs, there would have been some correlation with VWF levels. However, no such pattern was detected, with mean VWF levels remaining comparable between individuals grouped into reference, heterozygous 5'del-I and heterozygous 5'del-II genotypes (figure 4.10).

Further analysis was then undertaken after removing blood group O individuals from the data set (figure 4.11). This ultimately reduced the number of samples available for analysis, however, in non-O individuals there did appear to be a trend towards lower VWF:Ag levels in the 5'del-I group. The mean levels for this group were decreased by ~8% compared to the reference and 5'del-II groups. It is important to note however that this was not a significant difference and this was probably compounded by the low numbers in each group (reference: n= 110, 5'del-I: n=8, 5'del-II: n=9).

Taking the blood group O individuals alone, 5' del-I VWF: Ag levels were again decreased, in this case by ~4% compared to the reference group, but again this was not significant (figure 4.12).

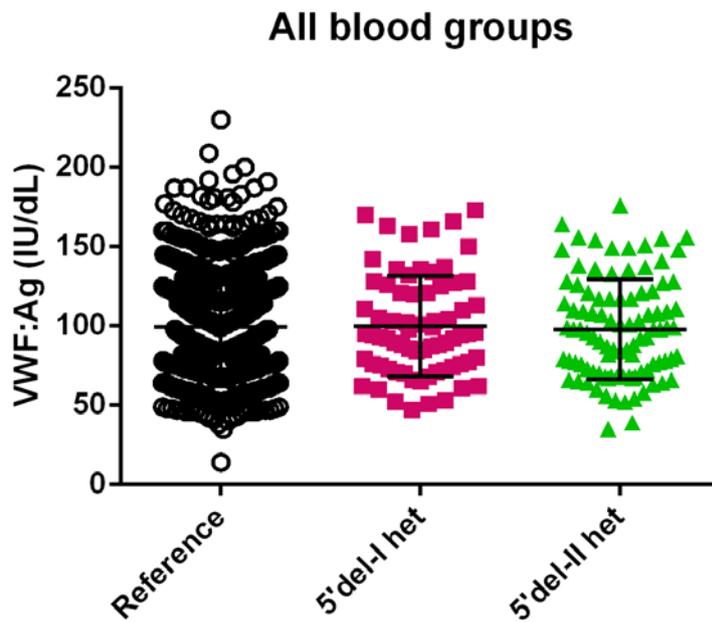


Figure 4.10 Correlation of 5' deletions with VWF:Ag levels in all blood group individuals. Mean VWF:Ag levels in each group and numbers of individuals for each group: Reference = 99.5 (n = 782), 5'del-I het = 99.8 (n = 71), 5' del-II het = 97.9 (n = 96). No significant correlation with mean VWF:Ag levels was detected in this group. Bars represent mean \pm SEM

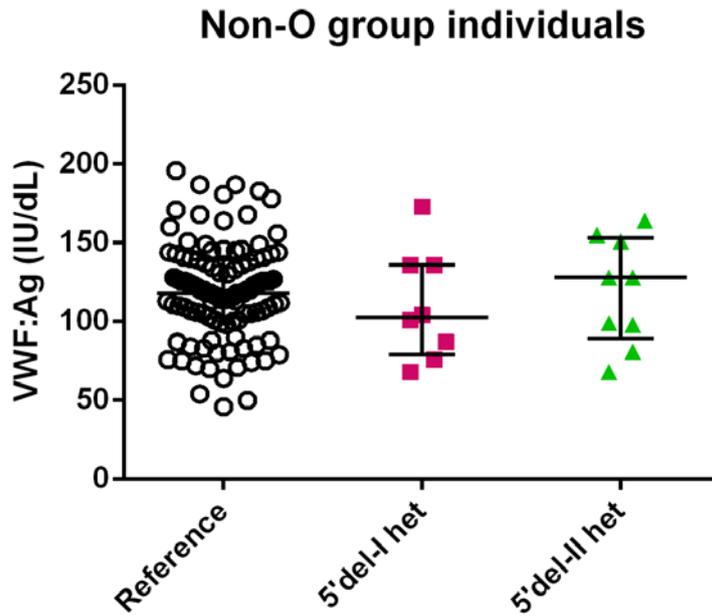


Figure 4.11 Correlation of 5' deletions with VWF:Ag levels in non-O blood group individuals. Mean VWF:Ag levels in each group and numbers of individuals for each group: Reference = 119 (n = 110), 5'del-I het = 110 (n = 8), 5' del-II het = 119 (n = 9). No significant correlation with mean VWF:Ag levels was detected in this group. Bars represent mean \pm SEM

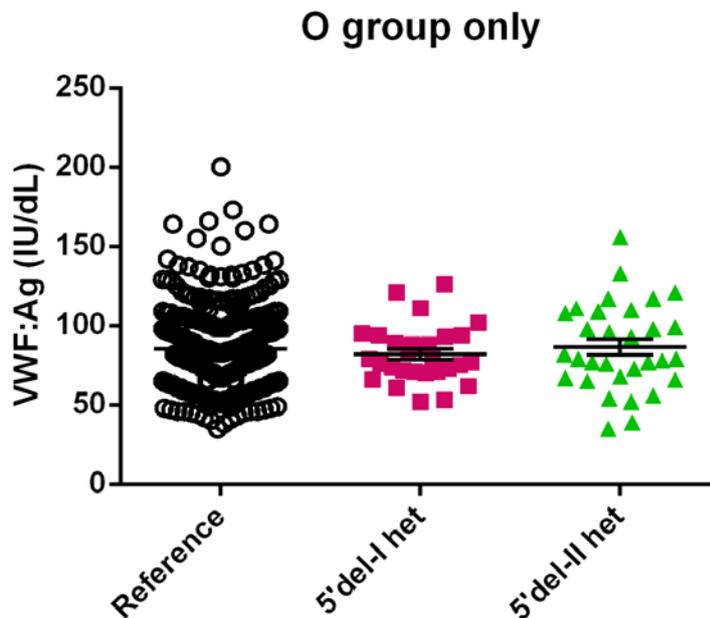


Figure 4.12 Correlation of 5' deletions with VWF:Ag levels in blood group O individuals. Mean VWF:Ag levels in each group and numbers of individuals for each group: Reference = 85.2 (n = 296), 5'del-I het = 82 (n = 28), 5' del-II het = 86.7 (n = 31). No significant correlation with mean VWF:Ag levels was detected in this group. Bars represent mean \pm SEM

4.5 Discussion

In this study, two non-coding, large deletion CNVs located in the upstream intergenic region of *VWF* were characterised. The deletion breakpoints were mapped and used to demonstrate the CNV deletion mechanism for each. This initial analysis showed that 5' del-I was a 6.64 Kb deletion and despite numerous repetitive elements in close proximity to the deletion breakpoints, the likely mechanism for deletion formation was microhomology mediated end-joining (MMEJ). Deletion 5' del-II was mapped as a 6.1 Kb deletion, where both 5' and 3' breakpoints were located within highly homologous AluY repetitive elements, suggesting Alu-mediated homologous recombination as the likely deletion mechanism. Alu-mediated HR appears to be a common mechanism around the *VWF* locus with numerous previous reports demonstrating the involvement of Alu repeats in deletion formation (Mohl et al., 2008, Sutherland et al., 2009).

The 80 Kb 5' flanking intergenic region of *VWF* contains numerous repetitive elements and is significantly enriched for Alu sequences (25% of the 5' intergenic sequence). Alu elements comprise ~10% of the human genome yet their functional role in genetic diversity and gene expression is still poorly understood. Repetitive elements such as SINES can give rise to functional non-coding sequences (Nishihara et al., 2006) and emerging evidence suggests that Alu elements may have some regulatory function. For example, recent work has demonstrated that Alu elements resemble enhancer sequences, in terms of active chromatin marks and specific histone methylation patterns. Interestingly, Alu elements were found to demonstrate higher levels of conservation when located in the upstream proximal region of genes. Furthermore, they appear able to influence the 3D genomic architecture through long range interactions (Su et al., 2014). The development of chromosome conformation capture techniques (Dekker et al., 2002) and its high-throughput derivative, Hi-C, has revealed the intricacies of genome organisation and architecture and its importance in gene regulation (Lieberman-Aiden et al., 2009). These DNA interactome maps have been utilised to demonstrate the involvement of repetitive elements such as Alu sequences, in the stabilisation of specific DNA-DNA interactions within the 3D architecture of metazoan genomes (Cournac et al., 2016). It is interesting to

speculate that the Alu elements in the *VWF* 5' intergenic region play a role in DNA looping and enhancer-promoter and/or silencer-promoter interactions. CNV such as the two deletions described here may disrupt looping and the associated distant acting effects of enhancer or silencing elements. Analysis of the AluY element adjacent to the *VWF* core promoter indicates that it shares 84% and 90% sequence homology with AluY elements at the 5' breakpoint regions of 5'del-I and 5'del-II respectively, highlighting the potential for Alu-mediated DNA looping and enhancer-promoter interactions.

Comparisons of Alu elements to shadow enhancers have also been made due to their similarity in distances from TSS in the region of 10-20 Kb (Su et al., 2014). These secondary enhancers are believed to confer some level of redundancy in enhancers, thus facilitating the evolution of these cis-regulatory elements without risk of disruption to their core enhancer functions (Hong et al., 2008). This may explain why mutation or loss of an enhancer does not always produce a clear phenotype. It could be hypothesised that the two enhancer regions identified in this study are secondary enhancers for *VWF* as the direct effect upon levels does not appear to be significant.

Precise breakpoint mapping also enabled a detailed *in silico* analysis of these intergenic regions, showing that both appear to possess regulatory potential. Recent data provided by the ENCODE project suggests that these regions could encompass active enhancers involved in *VWF* transcriptional regulation. Evidence for this hypothesis is based on the findings that these regions had HUVEC specific DNase hypersensitivity, transcription factor binding sites and key histone modifications, indicative of an active chromatin state. Significant regions of vertebrate conservation were also observed. This analysis supports previous *in silico* data from the Sheffield Haemostasis Research Group which suggests that 5' del-I encompasses a potential regulatory element with similar transcriptional activity to that of the *VWF* core promoter when studied in a luciferase reporter assay (Hampshire et al., 2009). Whether or not this reflects true enhancer activity in endothelial cells however, would require further work.

In the 5'del-I region, TF binding sites exist for NFYA, NFYB, MAFK, MAFF, FOS, JUN, CEBPB and MYC. As discussed previously, NFY TFs have been implicated

in the repression of the *VWF* promoter in non-endothelial cells, suggesting this region could act as a silencing element in some cell types. The Fos and Jun TFs have been shown to associate with GATA factors and could potentially form a complex that activates *VWF* promoter expression. In the 5'del-II region the TF sites were all located in a HUVEC specific DHS site in a region of active (H3K27ac and H3K4me1) chromatin. Binding sites for GATA2, Fos, Jun and POLR2A, possibly suggesting that some form of transcriptional activation complex could assemble in this region. The fact that this region is over 20 Kb from the core *VWF* promoter, may again suggest that a chromatin looping mechanism would be required for these regions to interact. It is interesting to note that the 1993 study by Ferreira and colleagues highlighted a 46 bp region upstream of the transcription start site that was essential for overcoming the downstream negative regulatory domain (Ferreira et al., 1993). This region resided within an Alu repeat that shares significant homology with numerous other Alu elements further upstream in the intergenic region, including those in close proximity to the 5' deletion CNV. It could be hypothesised that these Alu elements play a role in distant enhancer element interactions via facilitating chromatin looping interactions. This could be of significance also regarding the wider 5' intergenic region as other strong enhancer elements exist that may be involved in activating the *VWF* promoter.

Following up on the observation made in chapter 3 that the frequency of these deletions in a VWD cohort appeared significantly higher than that reported by the 1000 Genomes project, this study then tested the hypothesis that these deletions would be less frequent in a control cohort. In addition it was reasoned that these deletions may also correlate with *VWF* levels. In order to test these hypotheses, a multiplex PCR was developed and optimised, enabling the presence and zygosity of both 5' deletions to be analysed in a single reaction. This multiplex PCR deletion assay was used to screen a European healthy control (HC) cohort of 958 individuals. The frequencies of both deletions in this control cohort were similar to the VWD patient cohort analysed by aCGH. This would suggest that these deletions are not significantly more common in VWD patients than HC, with the caveat that the VWD cohort that was analysed was a relatively small sample size of 84 patients, all of whom were mutation negative. This would need to be

increased to make any confident conclusions about the frequency of these deletions in HCs versus patients. While no difference was observed between patients and controls in terms of the frequency of these two deletions, both groups showed a significantly higher frequency than those reported by the 1000 Genomes project.

In phase one of the 1000 Genomes project, the genomic coordinates mapping to 5'del-I show a frequency of deletions in this region of 37 in 1151 individuals (3.2%) (The 1000 Genomes Project Consortium, 2012). More recent data from phase three of this study shows a deletion frequency of 84 in 2504 (3.4%) (Sudmant et al., 2015b). Interestingly, a Dutch study of 767 individuals, reported a deletion frequency of just 0.13% (Boomsma et al., 2014), although this data was generated using Hi-Seq data, with CNV detection being performed based on strategies such as read-depth. Therefore it is possible that this region may be under reported for CNV in this study. A larger deletion of this region (~9.5 Kb) was also detected at a very low frequency by Cooper *et al*, in a large cohort of patients with intellectual disabilities and developmental delay. In this study just 3 deletions were reported in a sample size of 17421 (0.02%) (Cooper et al., 2011), using oligo aCGH and SNP arrays.

The phase one and phase three 1000 genomes project data for the 5' del-II region report deletion CNV frequencies of 21 out of 1151 (1.8%) and 95 out of 2504 (3.8%) respectively. In a study of CNVs in two groups of distinctly different ethnicities consisting of a Swiss European cohort and one from Sub-Saharan Africa, Vogler and colleagues reported a single larger (14 Kb) deletion of this region in just one individual (1/1109; 0.1%) (Vogler et al., 2010). It seems likely that some of these quite drastic differences in allele frequencies between this study and others reported in various databases may be due to different analysis techniques and whether sequencing approaches described in these other studies have sufficient coverage of the intergenic region in question. The repetitive nature of the DNA in this region may be a significant factor.

Samples from the 1000 Genomes project include a diverse range of ethnicities defined as European, Ad-Mixed American, African, East Asian and South Asian. The data from the European control group used in the present study suggests

that the deletion allele frequencies are significantly higher in Europeans than the rest of the world. Indeed, analysis of CNV data from the different populations reported in the 1000 Genomes database appears to support this hypothesis (see appendix 2). However, while this is an interesting observation, it is currently unclear whether these differences bare any significance to VWF level variation between populations.

Following analysis and comparison of VWF:Ag levels between the reference genotype, 5'del-I and 5'del-II, the data generated did not indicate any clear correlation. The results may indicate that these relatively common deletions, upstream of *VWF* do not impact significantly upon VWF levels in the general population despite the apparent loss of active chromatin sites that are characteristic of active enhancer elements. However, previous studies that have identified CNV deletions upstream of other genes, with comparable healthy control frequencies in the region of 10%, have been suggested to be risk factors for the disease (McCarroll et al., 2008, Wyszynski et al., 2016). Whether the two intergenic deletion CNV described in this study may be classed as risk factors for VWD is unclear. Further analysis of a larger cohort of both patients and controls may reveal interesting correlations with VWF levels. This highlights a major difficulty in associating a CNV with a complex and variable phenotype such as VWF level, as numerous other factors are involved. A more accurate and functionally relevant analysis in terms of *VWF* transcriptional regulation would be an analysis of mRNA levels in a large cohort of individuals with and without the deletion CNVs. Without such an analysis it is impossible to conclude whether these deletions influence transcriptional regulation.

Despite the proximity of these regions to VWF, the potential of long-distance chromatin interactions cannot be ruled out. A number of recent studies have highlighted the existence of distant regulatory elements playing a key role in gene expression and regulation. Therefore, although the 5' IR region of *VWF* is rich in regulatory active regions of chromatin, the target gene may not be *VWF*. It is interesting to consider that the 5' IR region of *VWF* may contain regulatory elements that act on distant genes that may be involved in regulating VWF levels. It is now understood that the human genome is organised into topologically associated domains (TADs) at the megabase scale (Dixon et al., 2012).

Disruption of these TADs via structural variations has been shown to have pathogenic consequences such as limb malformations, due to the disruption of long-range gene-enhancer regulatory interactions (Lupiáñez et al., 2015). It remains possible that regulatory elements for *VWF* exist elsewhere in the genome. As more data is generated on long distance chromatin interactions, in more cell types, novel *VWF* regulators may be revealed.

In summary, this study followed up the identification of two 5' intergenic deletion CNV, identified via aCGH. These were the only intergenic CNV detected during an analysis of a VWD cohort and therefore warranted further investigation. The breakpoints were mapped for each deletion, demonstrating MMEJ and Alu-mediated HR as likely CNV deletion mechanisms. The deleted regions were found to harbour significant HUVEC specific regulatory elements suggestive of a role in *VWF* regulation. A multiplex deletion PCR was developed to enable rapid screening of both deletions, highlighting similar frequencies in both patient and HC groups. However, both were significantly more common than the frequencies described in the 1000 Genomes project, suggesting a European population specific enrichment that is supported by the 1000 Genomes data. Although this study did not detect any significant effect of these deletions on *VWF*:Ag levels this does not mean that they do not have an effect on *VWF* expression. Analysis of mRNA from the different genotypes would be required to investigate any effect on gene expression further. While the identified 5' deletions may not be sufficient to cause a quantitative defect in *VWF*, they may well be a contributing factor, particularly in individuals whom may be compromised via another mutation for example.

Further *in silico* analysis of the 5' intergenic region of *VWF* highlighted numerous other putative enhancer sites and a significant enrichment for Alu repetitive elements. This enhancer rich, 5' intergenic region may harbour novel cis-acting *VWF* regulatory elements, which may also be potential sites for undiscovered mutations that contribute to VWD. Future studies should apply chromatin conformation capture techniques to study the interactions of these upstream elements with the *VWF* promoter region. Next generation sequencing may also reveal novel variants in these regions that impact upon *VWF* levels.

5 *In vitro* functional
characterisation of *VWF* exonic
deletion copy number variants
causing type 1 VWD

5.1 Introduction

In chapter 3 of this thesis, the use of aCGH for screening for CNV in *VWF* was presented in order to establish the prevalence of these genomic imbalances in VWD. This analysis and previous studies have highlighted the contribution of exonic CNV to the VWD mutation spectrum, however, in order to begin to understand and characterise the pathogenic mechanism of these variants in VWD it is desirable to perform expression studies. In the absence of endothelial cells from the patients carrying the CNV, heterologous expression systems are a viable alternative. HEK293 cells are a well-established heterologous system for *VWF* expression studies. When expressed in these cells, *VWF* undergoes constitutive secretion and is stored as HMW multimers in pseudo-Weibel-Palade bodies (WPB), which are indistinguishable from endothelial WPB (Michaux et al., 2003).

The biosynthesis and storage of *VWF* in WPB was described in chapter 1 and is an essential feature of *VWF* function. Without the densely ordered packaging of *VWF* tubules into WPB, the efficient release of HMW *VWF* strings is not possible. Disruption of tubulation impairs *VWF* string formation resulting in tangled, shorter strings with diminished platelet binding capacity and therefore reduced haemostatic functionality (Michaux et al., 2006). This biogenesis and packaging process may be disrupted by mutations or deletions in *VWF* that cause misfolding of the protein, leading to defective dimerisation and/or multimerisation and storage. The correct folding and dimerisation of *VWF* requires correct pairing and disulfide bond formation between the numerous cysteine residues that are present in the *VWF* protein sequence (Marti et al., 1987). This includes intrachain disulphide bond formation that is essential for *VWF* folding and also interchain disulphide bonds required in the formation of dimers and multimers during *VWF* maturation. Mutations of these residues therefore can disrupt the formation of dimers and multimers depending on the location of the mutated cysteine residue (Tjernberg et al., 2004). Mutations involving intrachain cysteine residues may disrupt the native structure of the *VWF* monomer leading to misfolding and ER retention, which in some cases may exert a dominant negative effect through the formation, retention and subsequent degradation of mutant-WT heterodimers by the proteasome (Eikenboom et al., 1996, Bodó et al., 2001). This ultimately

results in a severe quantitative deficiency in VWF. It is now understood that ER retention and defective intracellular storage and secretion from WPB is a common feature of quantitative VWF deficiencies in VWD (Wang et al., 2011), whereas disruption of residues involved in multimerisation prevent the formation of HMW multimers but do not cause a quantitative defect or disrupt WPB formation (Wang et al., 2012).

While many studies on quantitative VWF deficiencies have involved the characterisation of missense changes that cause substitution of key cysteine residues, the logic of disrupted disulphide bonds causing quantitative defects is applicable to type 1 VWD patients with large exonic CNV. Large in-frame deletions for example may result in the loss of numerous key cysteine residues, leading to aberrant folding of the mutant VWF and intracellular retention. Few studies have performed *in vitro* characterisation of deletion CNV and the effect of these large deletions on VWF storage in WPB is poorly understood.

The deletion of exons 4-5, identified in two patients during the aCGH analysis (chapter 3) in this thesis, has previously been shown to cause a severe reduction in VWF secretion in a dominant negative manner, presumably through the formation of heterodimers in the ER leading to retention (Sutherland et al., 2009). This deletion results in the loss of 104 amino acids including two cysteine residues (Cys159 and Cys162).

The deletion of exons 26-34, detected in a patient with dominant type 2 VWD (Bernardi et al., 1993) imparts a dominant negative reduction on VWF secretion (Casari et al., 2010). Interestingly, this type 2 mutant was able to form mutant-WT heterodimers and heterotetramers that were secreted via the Golgi, highlighting the fact that not all misfolded VWF is retained in the ER.

The exon 32-34 deletion has also been described in an Italian family, with type 1 VWD. *In vitro* analysis of this deletion mutant suggested that there was a significant reduction in VWF secretion and this was associated with increased intracellular lysate VWF levels compared to WT (Daidone et al., 2015). This study also provided evidence to suggest that the deletion caused reduced collagen binding and resistance to ADAMTS13. Interestingly the resistance to ADAMTS13 in this case was found to be due to an altered folding state in VWF brought about

by the deletion. This misfolded structure led to impaired ADAMTS13 binding which was reversible under denaturing conditions. In this study no microscopic characterisation of WPB formation was performed so the effect of this deletion on VWF storage is unclear.

Finally a recent study has described aberrant splicing as a cause of the exon 33 and 33-34 deletions. This study suggested that the pathogenic reduction in VWF levels in these patients was due to intracellular retention of the mutant VWF (Hawke et al., 2016). Patient BOEC analysis demonstrated fewer WPB with a rounder morphology, however it is important to consider that this patient was expressing three different splice forms of VWF in this case.

Prior to the aCGH analysis described in chapter 3, previous studies in the Sheffield Haemostasis research group utilised MLPA to screen for exonic CNV in *VWF* in mutation negative type 1 VWD patients recruited to the EU-MCMDM-1 VWD study. These analyses uncovered three large in-frame deletions in three separate families, that were later generated via site directed mutagenesis for in vitro expression analyses. The deletions included a deletion of exon 3 (rVWF3del), a deletion of exons 32-34 (rVWF32-34del) and a deletion of exons 33-34 (rVWF33-34del) (Hampshire et al., 2010, Cartwright et al., 2013).

The exon 3 deletion results in the loss of 55 amino acid residues from the D1 domain (p.Thr20_Gly74del). The IC of this family presented with a VWF:Ag level of 31 IU/dL, bleeding score of 10 and a normal multimer profile. The exon 32-34 deletion causes the loss of 128 residues, including nine cysteines and spans the A3, D4N and VWD4 domains of VWF (p.Val1820_Cys1948delinsS). The IC in this family had VWF:Ag levels of 12 IU/dL, a bleeding score of 12 and an abnormal multimer pattern. The third deletion of exons 33-34 resulted in the loss of 74 residues, including 8 cysteines and spans the D4N – VWD4 domains (p.G1874_C1948del). For this deletion the IC had a VWF:Ag level of 32 IU/dL, a bleeding score of 4 and an abnormal multimer profile. The full phenotypic data for IC and AFM of all three families are shown in table 5.1. Full family trees, phenotypic data and multimer profiles are presented in appendices 3, 4 and 5.

Table 5.1 Phenotypic data of IC and AFM with exonic deletion CNV

Patient	Status	CNV	Bleeding Score	VWF:Ag (IU/dL)	FVIII:C (IU/dL)	VWF:RCo (IU/dL)	VWF:CB (IU/dL)	RCo/Ag	CB/Ag	VWFpp/V WF:Ag	Blood group	Multimer Profile
P9F11 II:1	Index Case	Del Ex3	10	31	5	20	31	0.65	1	1.5	O/O	normal
P9F11 I:1	Affected	Del Ex3	10	35	21	28	25	0.8	0.71	1.6	O/A	normal
P9F11:II:5	Affected	Del Ex3	6	-	16	-	-	-	-	ND	A/A	NT
P9F3 I:2	Index Case	Del Ex32-34	6	12	17	11	12	0.92	1	4.8	O/A	abnormal
P9F3 I:3	Affected	Del Ex32-34	7	19	34	11	-	0.58	-	3.4	O/A	abnormal
P9F3 II:1	Affected	Del Ex32-34	5	12	22	8	-	0.67	-	4.1	A/A	abnormal
P6F1 II:1	Index Case	Del Ex33-34	4	32	25	23	15	0.72	0.47	4.1	O/O	abnormal
P6F1 III:2	Affected	Del Ex33-34	8	21	29	14	-	0.67	-	5.4	O/O	abnormal

- = data not available

Previous investigations in the group suggested that these deletion mutants all caused reduced VWF secretion compared to WT. However, a number of technical factors were identified that may have had some effect on the results obtained. These factors included the suitability of the cells used, differences in expression vectors used for each deletion, ELISA analysis and interpretation of ELISA data and the quality of the fluorescent microscopy images. These points are discussed in more detail below.

Firstly, previous experiments were performed using a batch of HEK293 cells that appeared morphologically different to healthy HEK293 cells (Cartwright, 2013). No information was available as to the number of passages of these cells and initial *in vitro* expression studies in these cells suggested that they were unreliable in their ability to produce pseudo-WPB. Other abnormal characteristics included a small, rounded morphology and a tendency to grow in clumps rather than an even mono-layer.

Secondly, two of the three deletion mutants were generated in the pCI-neo-VWF vector while the rVWF32-34del had been generated in pcDNA3.1. It was noticed that when WT VWF was expressed from the pcDNA3.1 vector in HEK293 cells, they failed to produce pseudo WPB-like structures as efficiently as those transfected with pCI-neo. Clearly this is an important consideration when comparing the effect of a mutant on WPB biogenesis to that of WT. To perform such a comparison therefore it is desirable for all mutants to be generated in the same vector to avoid any variation in the vector itself.

Thirdly, the normalisation of the ELISA data had been carried out using what transpired to be a false signal from the plate reader, thus introducing potential errors into the ELISA results. In addition, the cell lysis buffer used to prepare lysates for ELISA analysis had not been supplemented with protease inhibitors. This is likely to have resulted in the degradation of the target protein during cell lysis, therefore affecting the result obtained. These previous analyses had not taken into account total VWF production by each mutant (i.e. VWF in supernatant plus VWF in cell lysates) and this is a key consideration in order for meaningful interpretation of the data.

The final issue concerned the fluorescent microscopy analyses. Previous imaging performed for these experiments was of low resolution and quality, making interpretation of the results difficult as only subtle differences were observed between mutant and WT transfections, an issue that was probably compounded by the batch of cells that was used.

In summary, although these deletions were described previously, key questions regarding their effect on VWF biosynthesis, storage and secretion remain unanswered.

5.1.1 Aims

Taking these points into consideration, the aims of this study therefore were two-fold:

1. Generate the rVWF32-34del deletion in the pCIneo vector and perform *in vitro* expression of the three deletion CNV mutants in HEK293 cells. Experiments were designed to assay secreted and intracellular VWF levels via ELISA, to fully characterise the pathogenic mechanisms of these deletions that lead to reduced VWF levels. These investigations sought to take into account total VWF:Ag produced, the proportion of this VWF that was secreted into the media and the proportion of VWF that was retained in the cells. From these two measurements, the secretion ratio for each mutant could be calculated and compared to WT. This enabled two pathogenic mechanisms to be explored: reduction in total VWF levels (biosynthesis) and a reduction in the supernatant/lysate ratio (secretion).
2. To support the findings of the ELISA analysis, this investigation also aimed to provide high resolution fluorescent microscopy images to enable assessment of VWF intracellular localisation and the formation and morphology of WPB where present.

5.2 Results

5.2.1 Optimisation of pseudo-WPB production in HEK293 cells

The need to optimise the WPB production in HEK293 cells arose from the observation that previous work in the group had struggled to show pseudo-WPB formation in WT transfections. Stage one of the optimisation therefore, was to repeat these experiments using the wide-field fluorescent microscope. After staining for VWF, α -tubulin and DAPI, the imaging results demonstrated improved contrast and resolution over previous microscopy. However, cells that were producing WPB were almost undetectable. Instead the VWF staining was more diffuse throughout the cell, characteristic of ER localisation (figure 5.1 A).

It was reasoned that these results were due to the cells only having 48 hours post transfection to start to synthesise and package VWF into WPB. Therefore in the next round of transfections the post-transfection culture duration was increased to 72 hours.

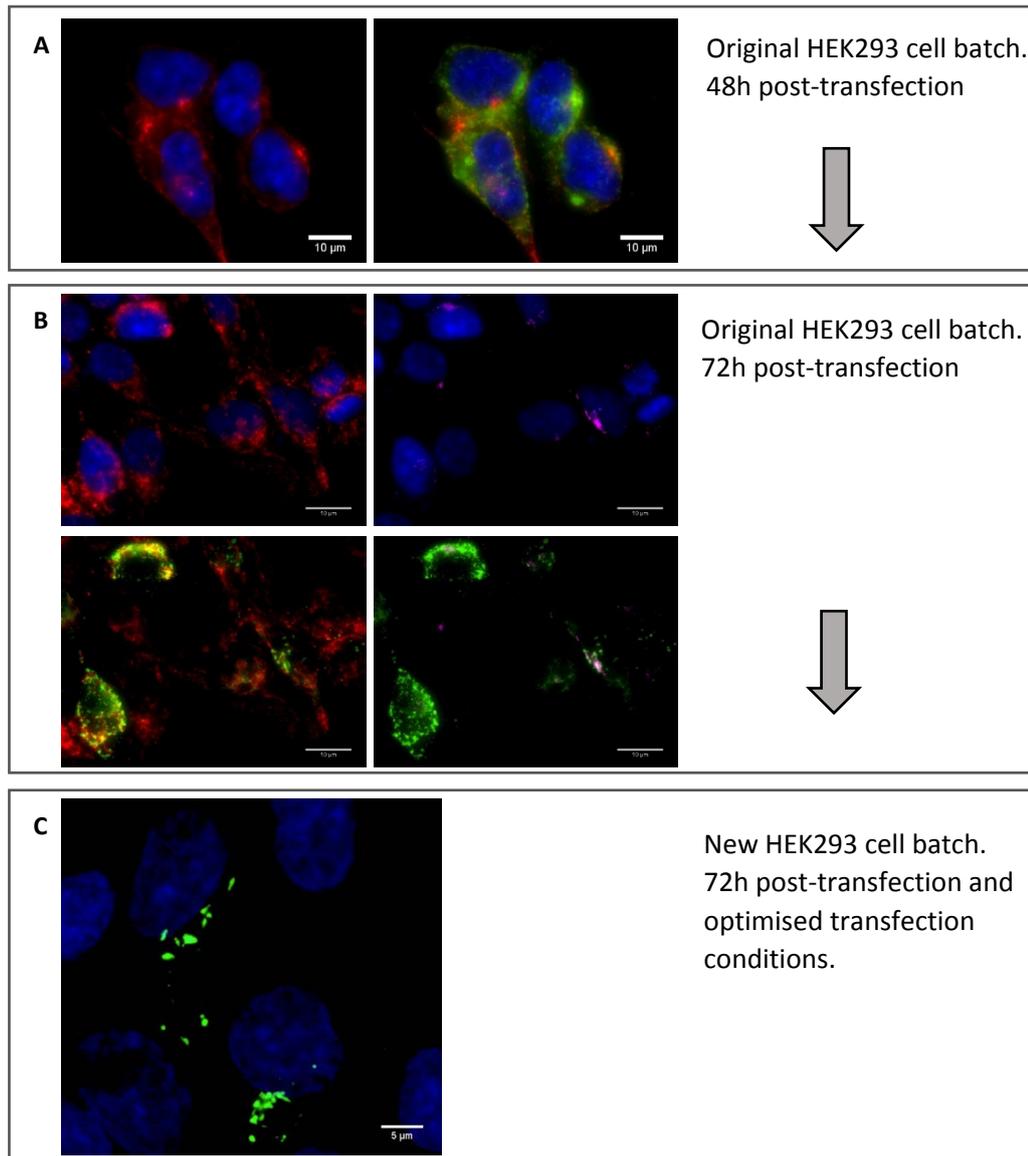


Figure 5.1 Optimisation of pseudo-WPB formation in HEK293 cells.

A. Original HEK293 cells stained after 48h with DAPI (blue), anti- α -Tubulin (red) and anti-VWF (green). WPB was absent. B. The same cells stained after 72h, using markers for the ER (anti-Calnexin; red) and TGN (anti-TGN46; magenta). Pseudo-WPB formation was visible in a small proportion of cells, although ER localised VWF was predominant. C. New HEK293 cells, transfected using an optimised transfection mix and imaged after 72 h. Numerous and more elongated pseudo-WPB formation was observed.

This extended culture period appears to be a commonly used culture time by other groups that have successfully imaged pseudo-WPB formation in HEK293 cells (Wang et al., 2011). This extended culture period brought immediate improvements with the formation of pseudo-WPB occurring in a small percentage of cells (figure 5.1 B). Combined with co-staining for the ER marker Calnexin and the trans-Golgi network marker TGN46, this initial result showed that a small population of cells were beginning to produce pseudo-WPB that were localised to the TGN, while a significant proportion of cells showed no WPB with the VWF signal localised with the ER marker. This observation led to the formation of a number of hypotheses on the cause of this issue and suggestions for possible solutions. These included:

- i. The HEK293 cells used for these experiments were beyond their optimum passage number for transfections and WPB biosynthesis.
- ii. The cell confluency may influence WPB formation as most cells observed to be making WPB had cell-cell contact.
- iii. The transfection efficiency was sub-optimal and/or the DNA concentration used for transfection was sub-optimal.

Each of these points was addressed in a series of optimisation experiments. A new batch of HEK293 cells was obtained from the research group of Dr. Kurt de Vos (University of Sheffield). These cells were originally obtained from the American Type Culture Collection (ATCC) and were supplied to the present study within passage seven. Under the light microscope, these had a distinctly different morphology to the previous HEK293s. This suggested that the cells used previously had been beyond their best and therefore were likely contributing to both poor transfection efficiencies and pseudo-WPB production.

The effect of confluency became apparent during the change to 72 hour culture post transfection. This extended period not only gave the cells more time to start expressing the transfected construct but also led to more cell growth and increased confluency. Therefore the improvements in WPB biogenesis observed after 72h is likely a combination of both factors. Nevertheless, experiments were repeated using a range of cell seeding densities in 6-well culture plates. These ranged from 150 k to 200 k cells per well. It was found that higher seeding densities improved confluency and ultimately led to improvements in WPB

formation. However, cell seeding densities over 200 k resulted in some over-confluent slides by 72 hours and therefore compromised imaging.

Previous protocols within the group had deviated from the recommended transfection protocol outlined in the Promega user guide by making up transfection reagent mixes in 2 ml of serum free media rather than the 100 μ l volume recommended. Presumably this was to reduce toxicity of the LTX reagent, as when cells were transfected using the recommended 100 μ l volume, significant cell death was observed. However, the larger volume of 2 ml was likely to be contributing to a reduced transfection efficiency. Therefore a range of transfection reagent volumes were tested over the range of 100 μ l, 500 μ l and 1ml. It was found that a volume of 500 μ l offered a sufficient compromise between cytotoxicity and transfection efficiency. The concentration of DNA was also investigated. While 2.5 μ g of the expression vector had been used in prior studies, it was thought this may be unnecessarily high and possibly contributing the number of cells showing ER localised VWF due to over expression. The hypothesis being that overloading the cells with a foreign DNA molecule which encoded a protein not normally expressed in this cell type was reducing the efficiency of VWF translation and packaging into pseudo-WPB. In order to test this, a range of DNA concentrations were transfected from 250ng/ml up to 1 μ g/ml final concentration. Based on these results and reference to protocols from other studies, a final DNA concentration of 700 ng/ml appeared to show improved results. Drastic differences were not observed between different DNA concentrations however, suggesting the improvement may be primarily down to the effect of DNA concentration on transfection efficiency.

Taken together, these optimisation steps using fresh HEK293 cells resulted in a significant improvement in cell viability, transfection efficiency and pseudo-WPB biogenesis (figure 5.1 C). This provided a heterologous system for WT and mutant VWF expression that could be used to compare differences with more confidence.

5.2.3 *In vitro* characterisation: ELISA analysis of VWF:Ag

In order to assess the effect of each mutation on total, secreted and intracellularly retained VWF levels, WT and mutant constructs were transiently transfected separately and in 1:1 ratio co-transfections, into HEK293 cells. After 24 h the media was replaced and the cells were cultured for a further 48 h prior to the supernatant being harvested for ELISA. Cells were washed twice with PBS before they were lysed with 1x passive lysis buffer containing a protease inhibitor cocktail. Quantification of VWF:Ag was performed via ELISA of both supernatants and cell lysates. Each mutant was transfected in triplicate and in three repeat experiments.

5.2.3.1 Reduced biosynthesis as a disease mechanism

ELISA analysis and quantification of supernatants and lysates was used to calculate the total VWF:Ag present (i.e. total VWF:Ag = VWF:Ag in supernatant + VWF:Ag in cell lysates) (figure 5.3 A). This initial analysis in rVWF3del demonstrated that in comparison to WT transfections, both heterozygous and homozygous transfections resulted in a significant decrease in total VWF:Ag. Heterozygous rVWF3del resulted in ~60% reduction compared to WT ($p < 0.0001$), while the homozygous form resulted in over a 90% reduction ($p < 0.0001$). This was a key observation that was necessary before making any conclusions about secretion or intracellular retention of the mutant by comparison to WT levels. A valid comparison cannot be made regarding secreted VWF of the mutant compared to WT without first ascertaining what percentage of the total is detected in cell lysates and what percentage is detected in supernatants. For example a direct comparison as a percentage of WT of the rVWF3del homozygous transfection would not indicate an increase in intracellular retention as the intracellular levels were similar to WT. However taken as a percentage of total VWF:Ag in that sample revealed that in comparison to the WT, a significantly higher proportion of the homozygous rVWF3del was found in the cell lysates.

Similar to rVWF3del, rVWF32-34del resulted in a significant decrease in total VWF:Ag levels (figure 5.3 A). In the heterozygous transfections, total VWF was reduced by 50% ($p = 0.0001$) and the homozygous transfection by 88% ($p =$

<0.0001). Again this demonstrated a severe and significant reduction in total VWF:Ag for this mutant in both cases.

The reduction in total VWF observed for rVWF33-34del was a milder yet still significant phenotype with the heterozygous transfections resulting in a 40% ($p=0.0011$) reduction while the homozygote resulted in a 53% ($p= <0.0001$) reduction in comparison to WT.

Despite the significant reductions in total VWF:Ag, the lysate levels of VWF were not significantly altered in comparison to WT levels ($p= >0.05$). The proportion of lysate to supernatant for each mutant compared to WT is shown in figure 5.3 B.

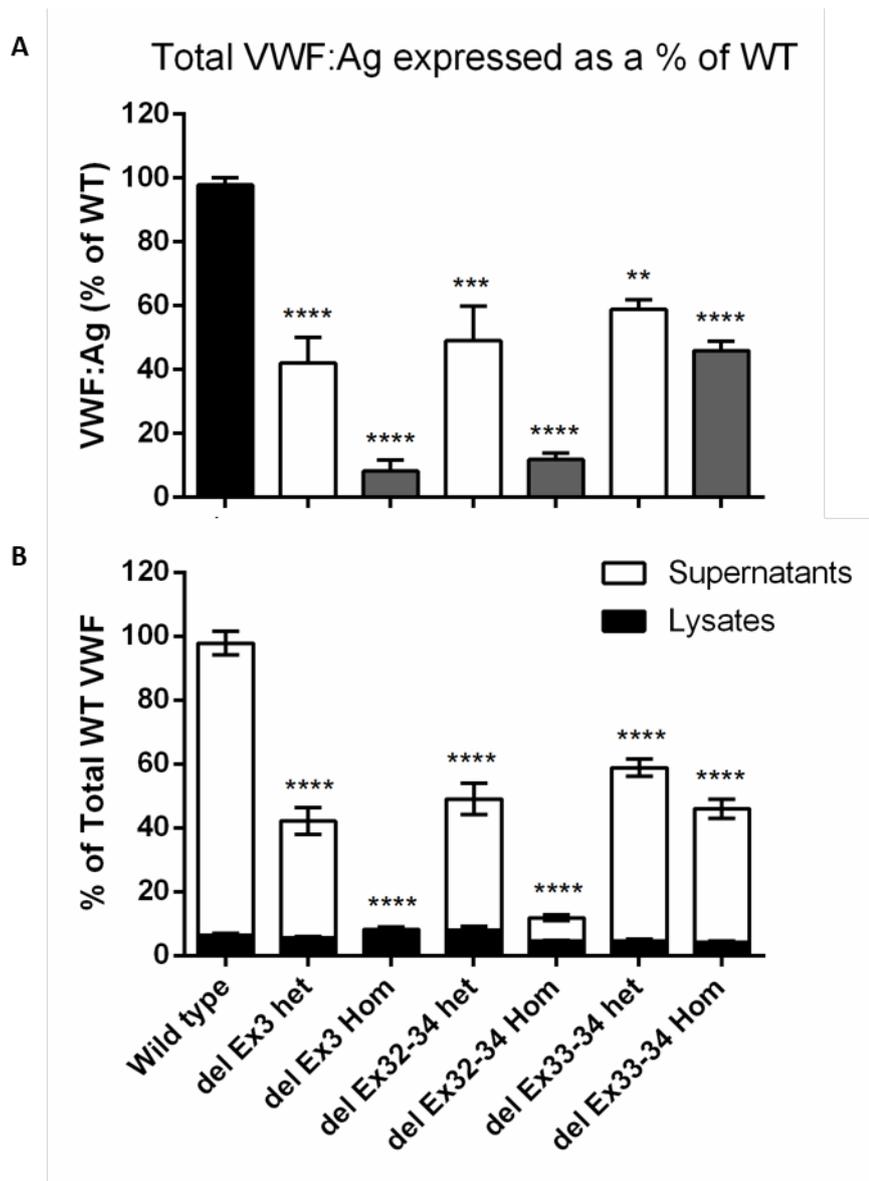


Figure 5.3 Quantification of total VWF:Ag in the deletion mutants as a percentage of WT.

A. Analysis of the total VWF:Ag synthesised by transiently transfected HEK293 cells revealed a significant reduction for all deletion mutants in both heterozygous and homozygous forms. Homozygous rVWF3del and rVWF32-34del resulted in the most severe reduction in antigen levels with 90% and 88% reductions respectively ($p < 0.0001$). **B.** Total VWF:Ag as a percentage of WT expressed as a proportion of lysate and supernatant. No significant differences were observed between mutant and WT lysate VWF:Ag levels. All VWF:Ag supernatant levels were significantly reduced compared to WT ($p < 0.0001$).

5.2.3.2 Reduced secretion as a disease mechanism

The secretion ratio was calculated by dividing the VWF:Ag in the supernatant by that in the cell lysates. Each mutant was normalised to WT, with the WT secretion ratio given to be 1.0 (see figure 5.4). The heterozygous rVWF3del had a dramatic effect on secretion resulting in a ratio of 0.41, a significant reduction compared to WT ($p = <0.0001$). In the homozygous form, secretion was almost undetectable with a ratio of just 0.016, indicating a major defect in, or non-existent secretion. This reduced secretion ratio alone is insufficient to fully explain the VWF levels observed in the patient however, with levels significantly lower than 40% that of UFM.

Calculation of the secretion ratio for rVWF32-34del also demonstrated a severe defect in VWF secretion for this mutant. In heterozygous transfections a 63% ($p = <0.0001$) reduction in secretion was observed (secretion ratio of 0.36) whereas in the homozygous form secretion was reduced by 90% (ratio of 0.1) ($p = <0.0001$).

Interestingly, analysis of rVWF33-34del revealed that the reduction in total VWF was a more significant contribution to the reduced VWF levels than the secretion defect in this mutant. The secretion ratio was calculated as 0.8 for the heterozygote and 0.73 for the homozygote representing reductions of only 18% and 15% respectively. This demonstrated that the largest influence on VWF:Ag levels in the supernatant was the reduction in total VWF and not the reduction in secretion.

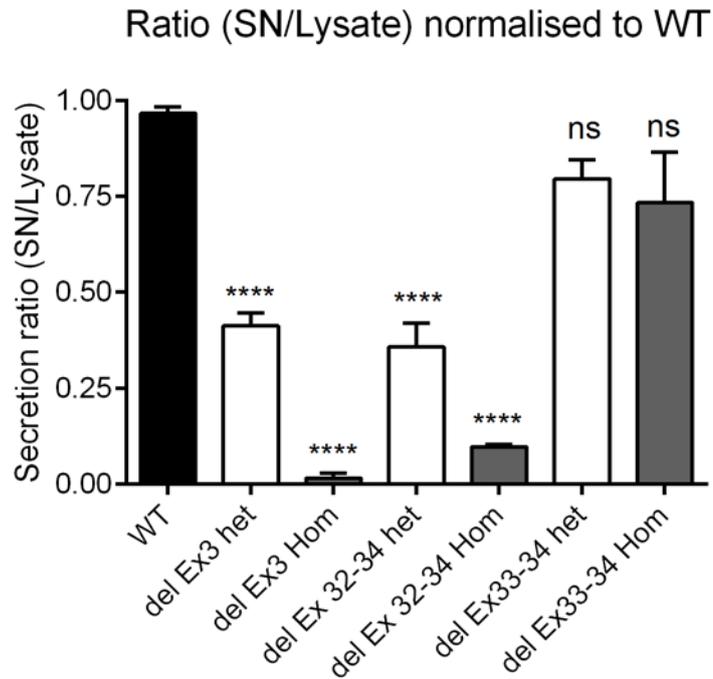


Figure 5.4 VWF secretion ratios for the deletion mutants normalised to WT (1.0).

The ratio of supernatant / lysate is presented for WT and each deletion mutant. Values were normalised to WT VWF secretion (taken as 1.0). Significant reductions in the secretion ratio were obtained for rVWF3del and rVWF32-34del in both heterozygous and homozygous forms ($p < 0.0001$). The reduction in secretion for rVWF33-34del was not significant ($p > 0.05$). Data were analysed by one-way ANOVA and Dunnett's post-test multiple comparison to WT. Error bars represent \pm SEM of three independent experiments in triplicate.

5.2.3.3 Effective secretion is a product of reduced total VWF and defective secretion

Analysis of the total VWF and secretion ratio data suggested that both reduced biosynthesis and reduced secretion (increased retention) were influencing the level of VWF:Ag in these mutants. Therefore, the reduction in levels observed in the patient, i.e. the disease phenotype, may be considered a function of the reduced VWF biosynthesis (or rapid VWF degradation) and the reduced secretion ratio (or increased retention ratio). It was found that by calculating this “effective secretion”, as a function of reduced total VWF and reduced secretion, a more accurate model of the patient phenotype is produced (figure 5.5). The mean effective secretion of het del Ex3 was 18.5% compared to the mean WT effective secretion of 95%. This represented a reduction in effective secretion of ~ 80%. Comparing this to the patient showed that a similar severe reduction was observed where the IC had a VWF:Ag level of 31 IU/dL and the UFM (parent) had a recorded level of 114 IU/dL, indicating a 73% reduction in levels as a result of the heterozygous del Ex3.

The severe reduction in total VWF and secretion ratio of rVWF32-34del, resulted in an effective secretion of 19% for the heterozygote and 1.2% for the homozygote. Comparison of these values to the P9F3 family showed that the IC had a VWF:Ag level of 12 IU/dL, while their unaffected offspring had a level of 122 IU/dL. In the patient therefore, the del Ex32-34 was causing a reduction in VWF levels of ~91% in comparison to the *in vitro* analysis which showed an 81% reduction in effective secretion. Clearly numerous factors impact upon the VWF levels *in vivo* which cannot be accounted for in an *in vitro* system such as that described here, but nevertheless, this analysis demonstrated that the severe reduction observed in the heterozygous patient was likely a combination of the impact of this deletion on total VWF levels, as a result of either rapid degradation or a significant reduction in biosynthesis and reduced secretion. These data, like that of del Ex3 suggested that in the homozygous mutant form the rVWF32-34del was resulting in an almost null phenotype.

For rVWF33-34del, neither the reduced total VWF nor the reduced secretion alone could account for the low VWF levels in the patient with this deletion. The del Ex33-34 index case had a VWF:Ag level of 32 IU/dL, while unaffected family

members ranged from 109 – 176 IU/dL, representing a reduction in levels in the region of 71% to 82%. The calculated *in vitro* effective secretion was 49% in the heterozygous state and 37% for the homozygous representing reduced secretion compared to WT of 51% and 63% respectively. This value more closely resembled the severity of this deletion in the P6F1 affected individuals, but still does not match the severity of the reduction in VWF levels.

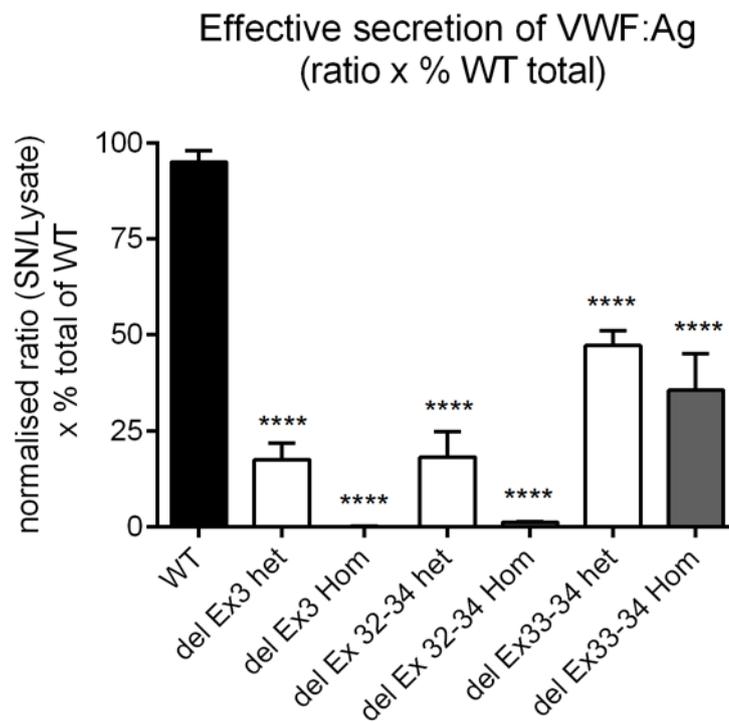


Figure 5.5 Effective secretion of VWF in WT and deletion mutants.

The effective secretion for each mutant was calculated by multiplying the secretion ratio by the total VWF production. Values were normalised to WT (100%). Significant reductions in effective secretion were observed for all deletion mutants in both heterozygous and homozygous forms ($p < 0.0001$) analysed by one-way ANOVA and Dunnett's post-test multiple comparison to WT. Error bars represent \pm SEM of three independent experiments in triplicate.

5.2.4 Widefield fluorescent imaging of VWF intracellular location and storage

Following optimisation of culture and transfection procedures for the reliable production of pseudo-WPB in HEK293 cells and ELISA analyses, the intracellular localisation of VWF was analysed for the three exonic deletion mutants. HEK293 cells were transiently transfected with mutant and/or wild type VWF. 72h post-transfection, cells were fixed and stained with fluorescent antibodies to the ER (anti-calnexin), trans-Golgi network (anti-TGN46) and VWF before imaging with a widefield fluorescent microscope. This allowed VWF to be monitored along the biosynthetic pathway from ER to Golgi before packaging into pseudo-WPB. Therefore any defects in WPB biogenesis due to the deletions would be highlighted.

When transfected into HEK293 cells, WT rVWF formed numerous pseudo-WPB often in close proximity to or emerging from the Golgi/TGN (figure 5.6 A-D). These were clearly visible as elongated VWF positive structures under widefield fluorescent microscopy. In these cells, ER localised VWF signal was minimal.

In the homozygous state, rVWF3del was unable to form pseudo-WPB and instead produced a diffuse staining pattern which appeared to co-localise with the ER marker calnexin (figure 5.6 E-H). This was highly suggestive of ER retention of the rVWF-delEx3. Over numerous repeated transfections and staining experiments, no WPB formation was observed, nor was any significant accumulation of VWF signal detectable at the TGN. Co-transfection with an equal quantity of rVWF-WT DNA to mimic the heterozygous state of the patient, restored some WPB formation, that could be observed clustered around the TGN, although these were relatively few in number compared to WT and some diffuse ER localised VWF remained (figure 5.7 E-H).

A similar result was observed for rVWF32-34del, with an ER localised VWF staining pattern and no clear pseudo-WPB formation. However, VWF positive structures were apparent (figure 5.6 I-L) although these did appear different to the ordered appearance of pseudo-WPB and were widespread throughout the cell with an apparent ER localisation suggesting that they may not have been storage organelles produced at the TGN. In co-transfection experiments, like

those for rVWF3del, the formation of pseudo-WPB was partially restored (figure 5.7 I-L).

In contrast to rVWF3del and rVWF32-34del, homozygous rVWF33-34del had a less dramatic microscopic phenotype, producing pseudo-WPB that appeared very similar to WT (figure 5.6 M-P). Although a number of smaller and rounder pseudo-WPB structures were present in this mutant, similar structures were present in the WT. Therefore it was unclear if the rVWF33-34del pseudo-WPB displayed any significant morphological differences compared to WT. In this mutant, the level of ER localised VWF appeared comparable to that observed in WT transfections.

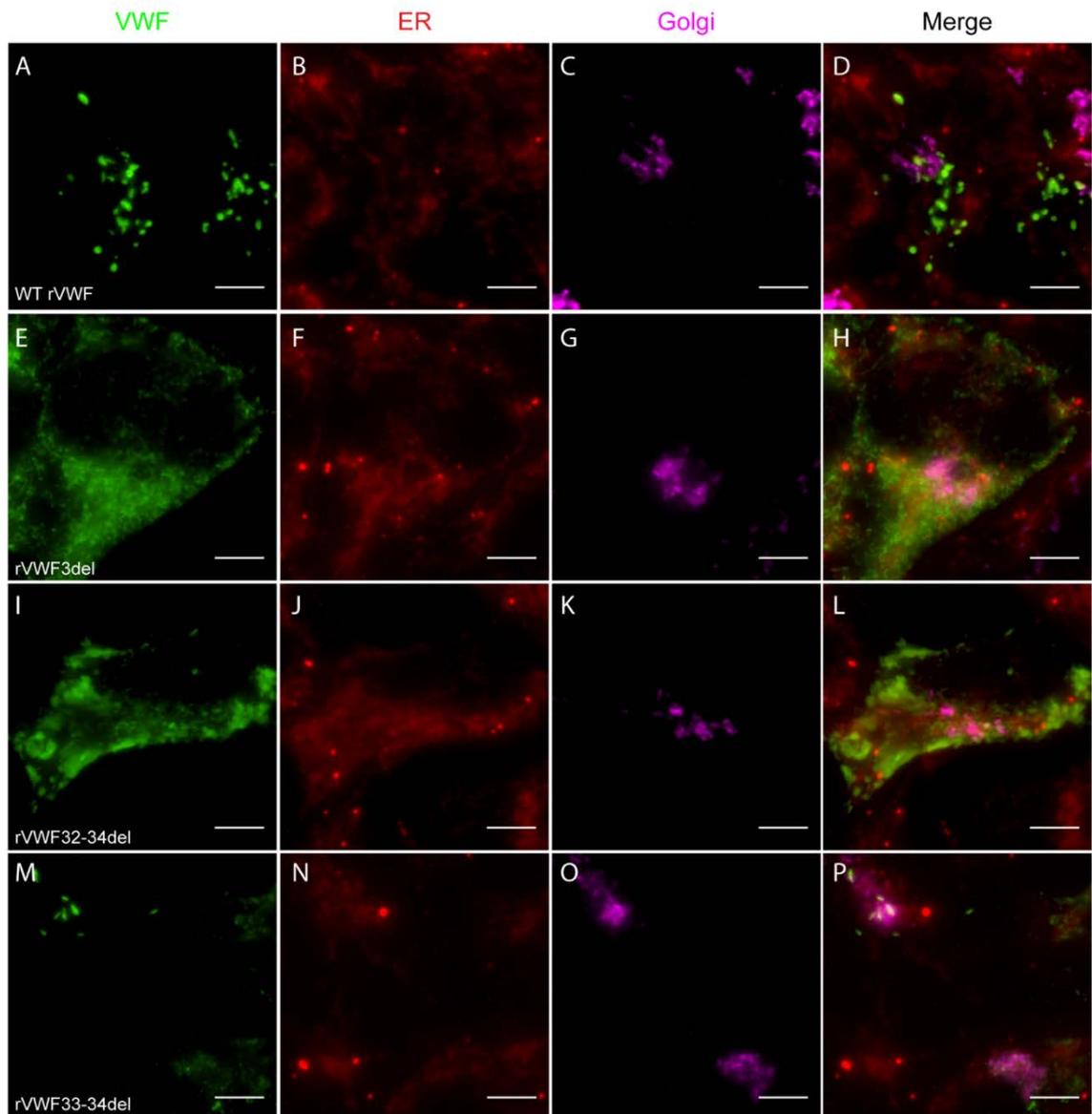


Figure 5.6 Widefield fluorescent microscopy of homozygous deletions

HEK293 cells were transfected with WT rVWF only (**A-D**) and either rVWF3del (**E-H**), rVWF32-34del (**I-L**) or rVWF33-34del (**M-P**). 72 hours post transfection, cells were fixed and stained with rabbit anti-human VWF, mouse anti-calnexin and sheep anti-TGN46. Primary antibodies were detected with secondary antibodies conjugated to alexafluors: goat anti-rabbit 488, goat anti-mouse 568 and donkey anti-sheep 647. Scale bars = 5 μ m

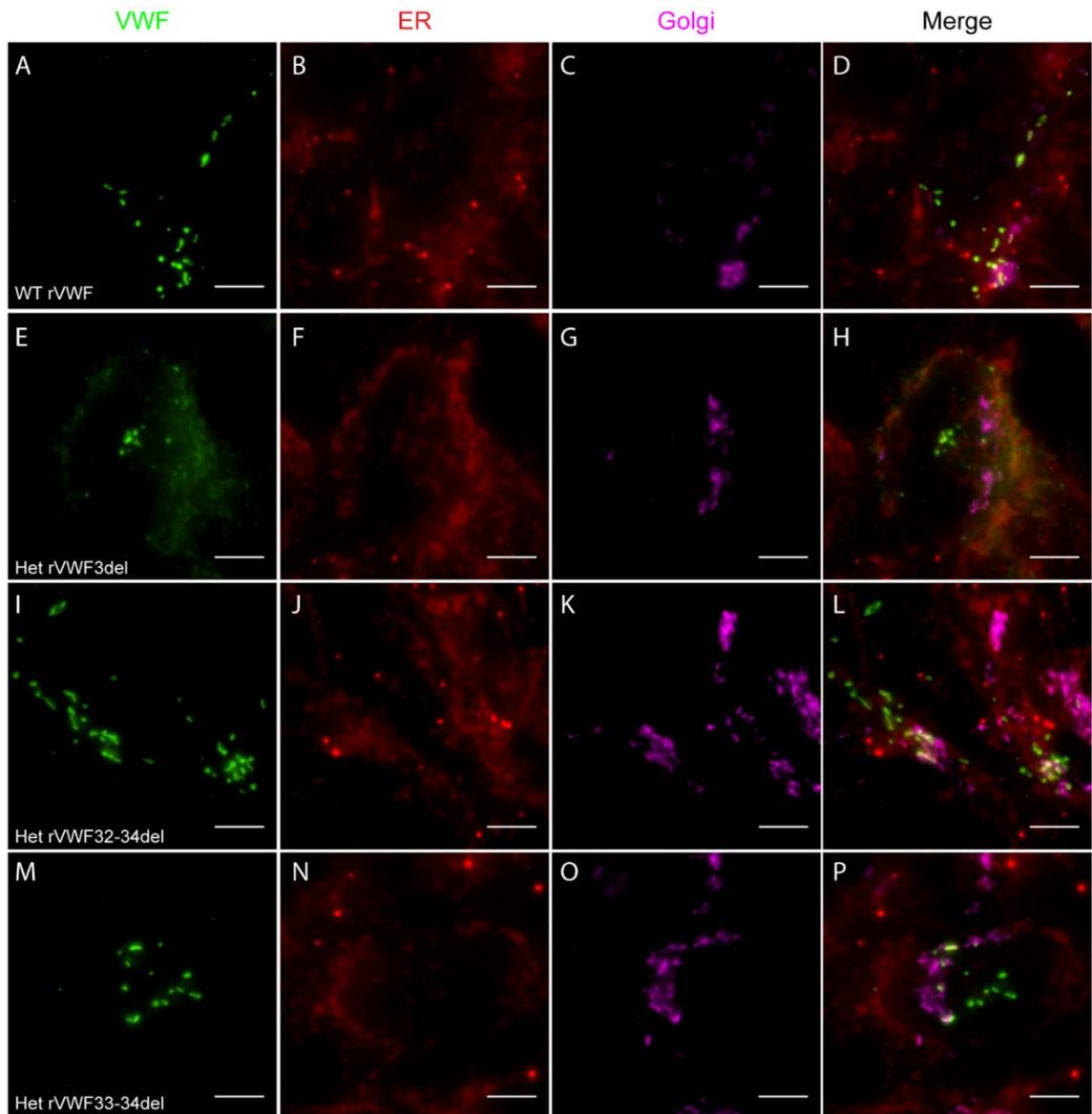


Figure 5.7 Widefield fluorescent microscopy of heterozygous deletions

HEK293 cells were transfected with WT rVWF only (**A-D**) and WT rVWF with either rVWF3del (**E-H**), rVWF32-34del (**I-L**) or rVWF33-34del (**M-P**) in a 1:1 ratio. 72 hours post transfection, cells were fixed and stained with rabbit anti-human VWF, mouse anti-calnexin and sheep anti-TGN46. Primary antibodies were detected with secondary antibodies conjugated to alexafluors: goat anti-rabbit 488, goat anti-mouse 568 and donkey anti-sheep 647. Scale bars = 5 μ m

5.2.5 Structured illumination microscopy of intracellular VWF and WPB

In order to obtain a more detailed understanding of the defects in VWF storage that each deletion mutant was causing, transfection experiments were repeated and cells were prepared for structured illumination microscopy (SIM). This super resolution technique uses spatially structured illumination to create a diffraction pattern on the sample. Modulation of the illuminating patterned light generates numerous Moiré fringes that contain the high resolution information and these can then be reconstructed to generate a super resolution image. This patterned illumination and reconstruction process, overcomes the diffraction limitations of conventional light microscopy, resulting in a two-fold resolution enhancement in both the lateral and axial dimensions (Gustafsson et al., 2008).

Prior to SIM imaging of the deletion mutants in HEK293 cells, a preliminary analysis was carried out on HUVECs and compared to WT VWF pseudo-WPB in HEK293 cells. This analysis provided a reference to compare the appearance of pseudo-WPB to WT endothelial WPB. The major differences between the two cell types in terms of WPB were that in HUVECs a greater number of WPB was routinely observed (>50) and many of these were over 2µm in length (figure 5.8 B). In comparison, HEK293 cells tended to produce fewer pseudo-WPB that usually appeared shorter than those in HUVECs, being around ~1µm or less (figure 5.8 A). The other clear distinction between the cells was the overall spread of the WPB. In HUVECs they could be found throughout the cell and often towards the cell extremities, with long WPB in parallel to the actin filaments. The distribution of pseudo-WPB in HEK293 cells was observable as a cluster on the apical side of the cell and close to the nucleus.

A clear observation of both pseudo-WPB and HUVEC WPB was the hollow staining appearance with bright outer membrane staining. This was even more apparent in the more elongated structures and likely reflects the limited antibody accessibility to those structures with densely packed VWF tubules. Overall, this comparison provided further evidence that HEK293 pseudo-WPB are very similar in morphology and structure to true WPB.

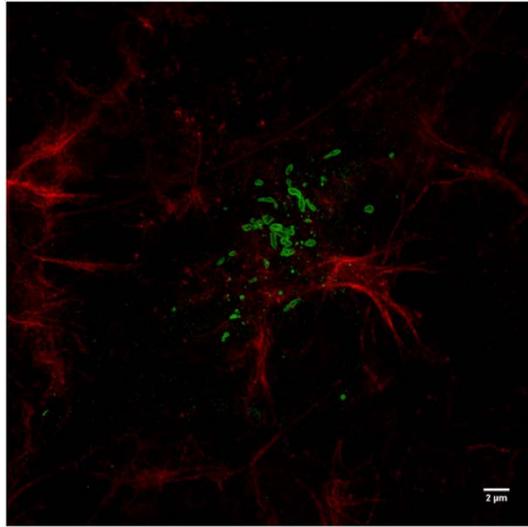
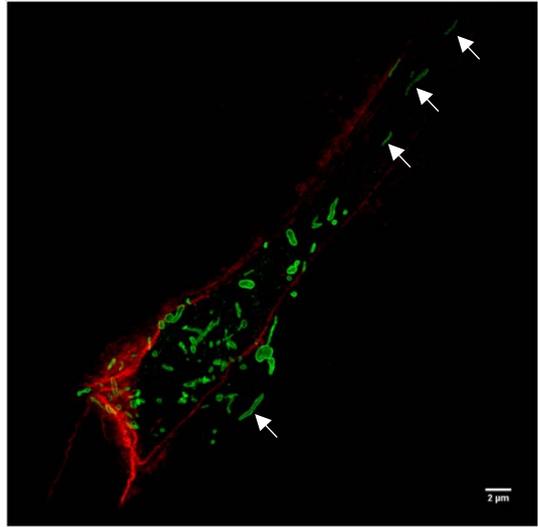
A**B**

Figure 5.8 Structured illumination microscopic comparison of WPB formation and morphology in HEK293 cells and HUVECs. **A.** Pseudo-WPB formed in HEK293 cells were generally shorter and fewer in number than those in HUVECs and tended to be more centrally clustered. **B.** HUVEC WPB were frequently observed as $>2\mu\text{m}$ in length, were more abundant and were distributed throughout the cell. Many HUVEC elongated WPB appeared to run parallel with the actin filaments toward s the periphery of the cell (B white arrows). Cells were stained with phalloidin to identify actin filaments (red) and rabbit anti-VWF with a goat ant-rabbit alexafluor 488 secondary antibody (green). Scale bars = $2\mu\text{m}$.

5.2.5.1 SIM imaging of pseudo-WPB formation in HEK293 cells transfected with deletion mutant VWF

Under SIM analysis, WT pseudo WPB were ~1-2 μm in length and had a hollow appearance with bright staining along the outer membrane (figure 5.9 A and E). Significant variation in shape and size of these structures was observable with some appearing as short round structures while others formed the classic elongated cigar shape. Some pseudo-WPB appeared to be in the process of maturing into elongated structures and contained swollen regions that may represent untubulated VWF (see white arrow in figure 5.9 E). Analysis of rVWF3del under SIM revealed a network-like structure of VWF with no formation of VWF positive puncta (figure 5.9 B). In comparison to WT the diffuse VWF staining was spread throughout the cell presumably reflecting its location, interspersed within the ER. Closer inspection revealed a spotted appearance of VWF in units in the nanometre size range that may represent accumulations of monomeric and/or dimeric forms of the mutant VWF (figure 5.9 F). Although rVWF32-34del appeared similar to rVWF3del when analysed on the widefield microscope, SIM analysis revealed a slightly different picture. The super resolution analysis appeared to show the formation of VWF positive puncta, more accurately described as VWF aggregates (figure 5.9 G). These were clearly distinct from pseudo-WPB in that they were disordered, random accumulations of VWF, without the characteristic hollow appearance observed in the WT rVWF pseudo-WPB and HUVEC WPB. Instead the VWF positive signal exhibited a bright uniform intensity across the entire structure, likely reflecting the disordered nature of the VWF within. Indeed it has recently been suggested that the relative intensity of VWF staining may be used as a measure of VWF tubulation and packaging within WPBs (Lopes da Silva et al., 2016). Disruption of VWF folding and the intra-organelle structure leads to a less compact state and provides an increased area of VWF to which the antibodies can bind. Using SIM the pseudo-WPB of rVWF33-34del were very similar to WT, although rounded structures were present with a uniform staining intensity (figure 5.9 H).

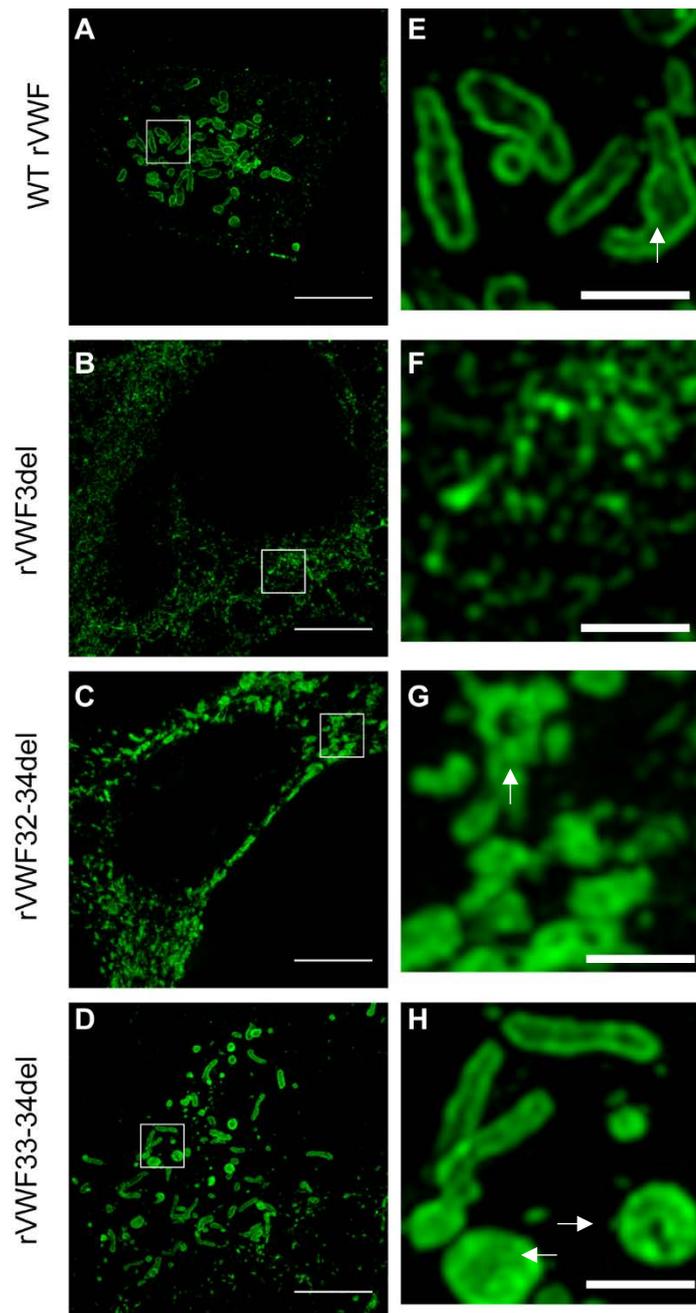


Figure 5.9 SIM imaging of pseudo WPB formation in WT and deletion mutants. Expression of WT rVWF produced numerous pseudo-WPB (A). On close inspection the elongated shape could be observed with bright outer membrane staining (E). The white arrow in E indicates a WPB containing untubulated VWF, possibly in the process of maturing. rVWF3del expression produced a diffuse, speckled network-like staining pattern (B). No large VWF positive structures were visible, suggesting the presence of monomeric and dimeric VWF only (F). rVWF32-34del also produced a more diffuse staining pattern although with apparent VWF positive structures (C), however, these displayed uniform staining intensity and non-uniform structure, possibly representing aggregates of VWF oligomers (white arrow) (G). rVWF33-34del was able to form pseudo-WPB like structures (D), however, numerous rounded structures were also visible (white arrows) (H). Cells were fixed and stained with rabbit anti-VWF primary, goat anti-rabbit alexa fluor 488 secondary (green). Scale bars = 5µm (A-D) and 1µm (E-H). E-H are higher magnifications of the region of interest in A-D.

5.2.6 Disruption of the signal peptide cleavage site as a potential mechanism for ER retention in VWF exon 3 deletion

The deletion of exon 3 results in the in-frame p.Thr20 _Gly74del. This loss of 55 amino acids deletes four amino acids from the C-terminus of the signal peptide and extends through into the D1 domain, deleting 51 aa including three cysteines. This deletion therefore, includes the signal peptide cleavage site. This region shows significant inter-species conservation around the cleavage site (figure 5.10 A), suggesting that loss of this specific amino acid motif may disrupt cleavage of the signal peptide.

The WT and deleted signal peptide sequence (first 40 residues of each) were analysed using the *in silico* signal peptide identification tool SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/> last accessed September 2016) (Petersen et al., 2011). In the WT sequence, the cleavage site was predicted to be after C22, in agreement with the published VWF SP cleavage site. This was indicated by a steep drop in the signal peptide score slope (S-score) after C22, a high cleavage score (C-score) and a peak in the combined cleavage score (Y-score) at A23 (figure 5.10 B). Analysis of the exon 3 deletion again showed a similarly high S-score, indicative of a true signal peptide which dropped off sharply but was staggered across residues 19-21. However, in this case the C-score and combined cleavage Y-score were both low with no clear indication of a cleavage site (figure 5.10 C). These data suggested that although the remaining residues within the SP would likely still target VWF to the ER, it was less certain that cleavage would take place. Furthermore, if cleavage did take place it was predicted to result in the loss of the first four amino acids of exon 4 (p.G74_Q77del), thus causing further disruption to the VWD1 domain in the propeptide.

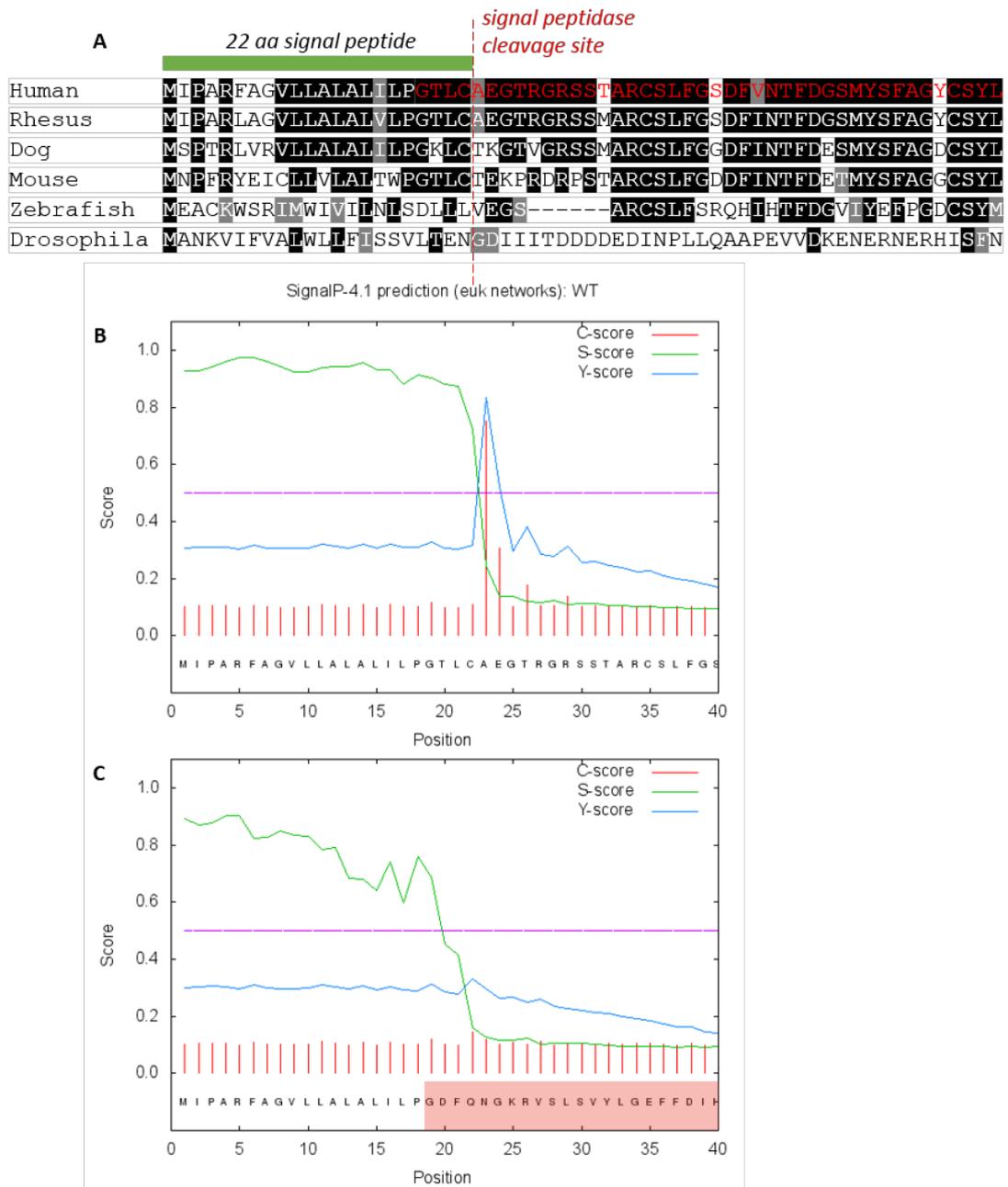


Figure 5.10 *In silico* analysis of the VWF signal peptide disruption in del Ex3.

A. Multiple sequence alignment showing the first 40 amino acids of VWF including the 22 aa signal peptide (green bar). Conserved residues are in shaded boxes. Note the conservation of the VLLALAL hydrophobic motif, C at -1, T at -3 and small residues such as A, T and G after the cleavage site. The red dashed line marks the SPase cleavage site and the red residues in the human alignment mark the start of the exon 3 deletion. **B.** Signal peptide and cleavage site prediction output from SignalP 4.1. Note the step uniform drop in the S-score curve after residue 22, and the high C-score and Y-score at the first residue of the mature peptide A23. **C.** The same analysis of the del Ex3 signal peptide region, showing a staggered S-score curve and weak C and Y scores, indicative of weak cleavage potential. The shaded red box indicates residues in exon 4.

5.3 Discussion

This study investigated the biosynthesis, storage and secretion of VWF for three exonic deletion CNV mutants; rVWF3del, rVWF32-34del and rVWF33-34del previously identified in a European type 1 VWD cohort. This analysis identified three major disruptions that occurred in these mutants to different extents, reflecting their disease phenotype. These pathogenic defects included reduced total VWF biosynthesis, a reduction in VWF secretion and impaired pseudo-WPB formation.

For all deletions a significant reduction in total VWF:Ag was observed. This reduction was most severe in homozygous rVWF3del and rVWF32-34del with both having around a 10-fold reduction in total VWF levels. The reduction in total VWF for homozygous rVWF33-34del was less severe with ~50% reduction. This is a significant finding and an important consideration when establishing whether a particular mutation has an effect on intracellular retention in comparison to WT. There are two potential explanations for this apparent reduction in total VWF:Ag. Either the cell is rapidly degrading the intracellular mutant protein so that the level remains at a low steady state or alternatively, the reduction in total VWF indicates a reduction in VWF biosynthesis which could be affected at the level of transcription or translation.

While no significant increases in VWF lysate levels were detected for the mutants in comparison to WT, the total VWF:Ag level should be taken into account. Therefore a more useful assessment of intracellular retention and reduced secretion, is to calculate the secretion ratio (VWF:Ag in supernatant / VWF:Ag in lysates) for each mutant and WT. This gives a more accurate picture of a secretion defect and associated intracellular retention. In the mutants studied here, significant reductions in secretion were observed for rVWF3del and rVWF32-34del. While the rVWF33-34del did result in a mild reduction in secretion, this was not significant.

Importantly, when comparing the reduced biosynthesis and reduced secretion levels to the severity of the defect in the patient phenotype, neither on its own could account for the reduction in levels observed in the patient. Therefore it was proposed that a more accurate picture of the *in vivo* situation is to take into

account both factors to give an effective secretion value. It is this value that indicates that rVWF3del and rVWF32-34del in the heterozygous state, may be acting in a dominant negative fashion with a reduction in effective secretion of >80% for both. Dominant negative effects on secretion have been demonstrated previously for the large exonic deletions of exons 4-5 (Sutherland et al., 2009) and exons 26-34 (Casari et al., 2010). Degradation of ER retained dimers, i.e. those that contain a mutant subunit, has been described as a potential mechanism for the dominant type 1 VWD reduction in VWF levels (Bodó et al., 2001). However, this hypothesis suggests that assuming random dimerization then only 25% of VWF dimers would leave the ER i.e. a 75% reduction in secreted VWF. In the present study heterozygous rVWF3del demonstrated only a 60% reduction in secretion but also a 60% reduction in total VWF levels. Furthermore, heterozygous rVWF33-34del resulted in a 65% reduction in secretion and a 50% reduction in total VWF. Type 1 VWD also frequently results in a greater than 75% reduction in VWF levels as observed in the present study for patient P9F3 II:1 whom has a ~90% reduction in VWF:Ag levels compared to UFM. Interestingly this patient was also heterozygous for the 5'del-I deletion discussed in chapter 4. Even so, there remain many other factors that contribute to *in vivo* VWF levels that cannot be recreated and modelled in an *in vitro* situation as described here.

This suggests that mechanisms other than mutant heterodimer retention are involved in the overall reduction in VWF and that reduced biosynthesis is an important factor. A reduction in total VWF biosynthesis is interesting and has been reported previously for the p.Arg1205His and p.Tyr1584Cys mutations, where expression in a range of cell types and in hydrodynamic injection of mice demonstrated reduced VWF biosynthesis of human and murine VWF (Pruss et al., 2011). Two mutations in the A3 domain of VWF (p.L1696R and p.P1824H) have also been shown to cause intracellular retention and reduced expression both *in vitro* and *in vivo*, with p.P1824H causing a three-fold reduction in total VWF levels. *In vivo* the reduced levels were found to be even more pronounced (Legendre et al., 2013).

What mechanisms this reduced biosynthesis may involve are unclear. It would appear that some negative feedback mechanism may be in action that is able to alter the synthesis of VWF in response to reduced secretion and increased

intracellular/ER levels. One possible explanation for this is the activity and output of the unfolded protein response (UPR), brought about by an increase in misfolded VWF within the ER. An increase in the load of unfolded or misfolded protein in the ER leads to the initiation of an ER stress response, whereby the cell activates a number of pathways aimed at reducing this load (Schröder and Kaufman, 2005). Misfolded proteins are detected in the ER by chaperones such as BiP, leading to activation of the UPR. This response involves increasing the folding capacity of the ER via upregulation of folding chaperones (Kozutsumi et al., 1988) and reducing the secretory protein load by down regulating translation via PERK mediated phosphorylation eIF2 α (Harding et al., 1999). ER-associated degradation (ERAD) of misfolded proteins is also upregulated (Travers et al., 2000), whereby misfolded proteins are retrotranslocated into the cytosol, where they are ubiquitinated and thus targeted to the proteasome for degradation.

The literature on ER stress in relation to VWF however, is relatively limited. VWF has been shown to interact with protein folding chaperones such as BiP and mutant VWF that is retained in the ER has been shown to undergo a prolonged interaction with ERp57 (also known as PDIA3) (Allen et al., 2000, Allen et al., 2001). Homocysteine causes VWF ER retention (Lentz and Sadler, 1993) via disruption of disulphide bond formation, leading to ER stress, growth arrest and changes in gene expression in HUVECs (Outinen et al., 1999). Outinen and colleagues observed that when cultured in the presence of homocysteine, HUVECs showed measurable cellular changes consistent with ER stress. Firstly, VWF from these cells could be immunoprecipitated with the ER chaperone BiP that binds misfolded proteins, and expression of transcription factors such as GADD153 (CHOP) and ATF-4 were increased. These are now known to be ER stress inducible factors.

Interestingly, further evidence that supports the idea of misfolded, ER retained VWF causing ER stress is the use of Tunicamycin as a positive control in ER stress induction experiments. Tunicamycin has been shown previously to have an endothelial specific effect on VWF glycosylation and dimerization, leading to retention of pro-VWF subunits in the ER (Wagner et al., 1986). Therefore a drug used to induce ER stress also causes VWF ER retention.

All this evidence points to a mechanism by which ER retention of misfolded VWF, induces ER stress and translational arrest leading to down regulation of protein synthesis. This may in part explain the apparent reduction in total VWF observed in these mutants. However, there is a further line of evidence that warrants some consideration. At the level of gene expression regulation of VWF, our understanding is limited as very few studies have investigated VWF mRNA levels in type 1 VWD patients. Interestingly, one study has gathered such data from type 1 VWD patient BOECs. This work demonstrated significantly reduced VWF mRNA in type 1 BOECs compared to HC and type 2 VWD patients (Starke et al., 2013). Some of these BOECs showing reduced VWF mRNA were derived from a patient heterozygous for the exon 4-5 deletion which is known to cause ER retention and reduced secretion in a dominant negative manner. Furthermore mRNA levels showed a linear correlation with VWF:Ag levels, indicative of VWF levels being linked to VWF gene transcription in these patients. The causative mutations in these patients did not account for this reduction in VWF mRNA. Other studies on type 1 VWD, that have also measured VWF mRNA are few and far between, with only one previous report demonstrating reduced VWF mRNA in endothelial cells from a type 1 patient (Ewenstein et al., 1990). It is reasonable therefore to hypothesise that reduced VWF levels in some quantitative defects are brought about in part at the level of gene expression regulation or rather mRNA production and/or stability.

An explanation for this observation may also be found in emerging evidence from the ER stress field. During ER stress, a second ER protein folding sensor, Ire1a is activated via autophosphorylation. At low levels of ER stress, Ire1a cleaves the mRNA of *XBP1* leading to an alternatively spliced transcript (*xbp1s*). This splice product encodes a transcription factor which increases expression of ER stress responsive genes such as folding chaperones, to help relieve some of the burden on the ER (Calton et al., 2002). Ire1a nuclease activity has also been shown to be active against other mRNAs and is able to degrade mRNAs that are targeted to the ER, often encoding secretory proteins that place a greater burden on the ER. This process is termed regulated Ire1 dependent decay (RIDD) (Hollien and Weissman, 2006, Hollien et al., 2009). These mRNAs share common Ire1a cleavage and structural motifs with *xbp1* mRNA. These motifs include a

conserved CUGCAG sequence where cleavage occurs after the second base and this sequence is flanked by self-complementary bases that are able to form a stem-loop (sl) structure with the CUGCAG cleavage sequence sitting in the unpaired loop structure (Oikawa et al., 2010). Initial *in silico* analysis in the latter part of the present study suggests that VWF mRNA contains seven CUGCAG Ire1a cleavage sequences. At least two of these also form within the loop of the conserved stem loop structure, raising the possibility that VWF mRNA may be a substrate for RIDD. This novel mechanism would link VWF mutations causing ER retention with ER stress, translational arrest and mRNA degradation.

In this study the *in vitro* expression results are supported by the microscopic characterisation of VWF storage and WPB formation. Both standard widefield fluorescent microscopy and super resolution SIM imaging techniques were employed to demonstrate the impact of exonic CNV on the intracellular storage of VWF. While previous studies have reported reduced VWF levels in association with defects in WPB formation in type 1 VWD due to missense mutations in different VWF domains (Wang et al., 2011, Castaman et al., 2012, Wang et al., 2012), few have explored storage impairment due to pathogenic CNV.

In WT VWF transfections, numerous elongated pseudo WPBs ~1-2 μm in length were observed (Fig. 5.9A and E). These storage organelles had a distinctive staining pattern, with SIM analysis revealing a brightly stained outer membrane, giving the pseudo WPB an almost hollow appearance. Interestingly this super resolution microscopy also revealed significant variation in the morphology of these structures, with some appearing to have more irregular proportions than the classic cigar shape. Presumably this hollow appearance seen in WT WPB is due to the tubulation and dense, ordered packaging of VWF into the WPB resulting in limited antibody access to the internal VWF epitopes (Geysen et al., 1987). The observation of irregular shaped WPB in WT cells may reflect the presence of untubulated material in the lumen of the immature WPB that has been delivered from the Golgi, prior to its tubulation (Mourik et al., 2015b). It is important to be aware of this heterogeneity in the WT WPB population, prior to assessing the effects of mutations or CNV on VWF maturation, storage and WPB morphology.

The lack of WPB formation in rVWF3del may have been expected given the essential role of the propeptide in VWF tubulation in WPB formation (Michaux et al., 2006). ER retention has been described previously for mutations within the VWF propeptide. The D1 and D2 domains contain key residues that are N-glycosylated and this is essential for the trafficking of VWF dimers to the Golgi (McKinnon et al., 2010). Furthermore, key CXXC motifs within D2 are thought to form a transient covalent bond with D3 in the ER prior to Golgi multimerisation (Purvis and Sadler, 2004). These key functional sites are disrupted in the previously reported del Ex4-5, which also results in ER retention of the mutant protein (Bowman, 2013). While these regions are not within the del Ex3, they are in close proximity within the propeptide and therefore it is possible that the conformational consequences of this exon deletion may disrupt the correct folding of these regions, thus perturbing their normal function.

However, a more likely factor in the ER retention of the exon 3 deletion is the disruption to the C-terminal of the signal peptide. Signal peptides (SP) such as that of preproVWF, are recognised by the signal recognition particle (SRP) as they emerge from the ribosome (Walter and Blobel, 1981). SP binding by SRP initiates elongation arrest in order to maintain a translation competent state while awaiting a vacant SRP receptor (SR) site on the ER membrane (Lakkaraju et al., 2008). Docking of the ribosome nascent chain-SRP complex (RNC-SRP) at the SR on the ER membrane causes the SRP and SR to dissociate and translation is resumed in a co-translational translocation into the ER lumen, during which the signal peptidase (SPase) cleaves the SP (Meyer et al., 1982). Signal peptides consist of three core regions or domains, the variable N domain with a net positive charge, an H or hydrophobic region (residues ~ -5 to -13) and a C region (-1 to -5) which is not variable in length and contains specific residues that define the cleavage site (von Heijne, 1985). The C region defines the cleavage site by specific residues at -1 and -3, with these usually being small uncharged residues such as Ala, Gly, Ser, Cys, Thr or Gln. Conversely residues with bulkier side chains are commonly found at the -2 position. Positions -4 and -5 are usually Gly and Pro residues and Pro is never found between -3 and +1 (Von Heijne, 1983). Insertion of a Pro residue at +1 next to the cleavage site has recently been shown to bind and inhibit the catalytic subunit of the SPase thus disrupting cleavage (Cui

et al., 2015). The C region of the signal peptide then is highly conserved and an essential component of the signal peptide that defines the correct cleavage site. The deletion of exon 3 described here results in deletion of four of the five C region residues, leaving Pro18 as the last residue in the SP sequence. However, as described this is incompatible with cleavage and therefore the site is shifted, resulting in a predicted weak cleavage site after Gln76. However this would leave a Gln residue at -1 and an Asp residue at -3, neither of which fit the convention. Furthermore, the +1 residue, i.e. the first in the mature peptide, would be Asn77. Asparagine residues are not commonly found at this +1 site. This evidence suggests that the exon 3 deletion mutant does not undergo signal peptide cleavage at the ER or at best if cleavage does occur it would be inefficient and would delete a further four residues from exon 4. The fate of the uncleaved del Ex3 prepro-VWF remains unclear, although it is likely to remain anchored to the ER membrane where it may disrupt the processing and targeting of other secretory proteins including WT VWF. A similar mechanism has been proposed in a form of central diabetes insipidus (CDI) where a mutation in the SP of vasopressin caused reduced cleavage efficiency (Ito et al., 1993). This could also result in a dominant-negative mechanism in addition to the formation of heterodimers. Heterodimer formation in the present study is presumed due to the presence of dimers in the *in vitro* multimer profiles for the homozygous transfections, suggesting that dimerisation can still take place. This may be expected given that no disruption occurs in the C-terminal region of the protein. Signal peptide mutations and ER retention have been reported previously in other pathologies including hypoparathyroidism (Datta et al., 2007), and more specifically, pathogenic mutations have been identified in the signal peptide cleavage sites of antithrombin (Daly et al., 1990), coagulation factor X (Racchi et al., 1993) and preproinsulin, causing neonatal diabetes (Støy et al., 2007). The only previous report of a mutation in the VWF signal peptide is Gly19Arg (Eikenboom et al., 2009). This is a substitution at the -4 position and therefore not in an essential cleavage residue position, although some effect on cleavage efficiency cannot be ruled out. However, it is thought that the likely effect of this mutation may be disruption of the donor splice site. Therefore the exon 3 deletion in the present study would be the first report of a signal peptide cleavage defect in VWF causing type 1 VWD.

It has been suggested that only the D1-D2-D'-D3-A1 domains are required for the formation of elongated WPB (Michaux et al., 2006), however a recent characterisation of 6 mutations within the A1-A3 domains, suggested that mutations in the A domains can also impair VWF storage (Groeneveld et al., 2014). Of interest to this study, p.V1822G within the A3 domain, resulted in significant ER retention. Reduced or lack of response to DDAVP is associated with mutations in the D3 – A3 domains of VWF in type 1 VWD, suggesting that residues or motifs within these regions are essential for the biogenesis of functional WPB (Castaman et al., 2008). The fact that rVWF32-34del extends into the A3 domain may partly explain the differences observed *in vitro* in terms of loss of HMW multimers and lack of WPB formation between this mutant and rVWF33-34del. It has been speculated that the A1 domain plays some role in stabilising VWF tubules (Huang et al., 2008) and perhaps A3 performs some similar structural role. It is also possible that loss of the residues within the A3 domain leads to severe misfolding and accumulation of mutant VWF within the ER. Presumably this increased accumulation in the ER of rVWF32-34del leads to the aggregations observed under SIM. In some instances, large VWF positive compartments or vacuoles were observed, possibly reflecting the dilated nature of the ER in these mutants. Dilation of the ER is a common observation in mutant proteins that are retained in the ER, including a number of VWF mutants that disrupt cysteine residues involved in intrachain disulphide bond formation (Wang et al., 2012). The difference in severity observed between rVWF32-34del and rVWF33-34del is quite striking. The additional cysteine deleted in rVWF32-34del is Cys1872 and loss of this residue may be the difference, possibly leading to a greater lack of stability and greater tendency to misfold. The report from Daidone et al suggested that this deletion mutant was resistant to ADAMTS-13 due to its abnormal folded state which was overcome in denaturing conditions (Daidone et al., 2015). This aberrant folding may be reflected in the aggregations observed in the SIM analysis.

The fact that rVWF33-34del formed similar if not identical pseudo-WPB as WT VWF was unsurprising given the relatively mild effect this deletion has on VWF multimers and secretion *in vitro*. Furthermore, patients with this deletion have higher levels than those with the other two deletions identified in this study. The

presence of normal pseudo-WPB in this mutant fits with the *in vitro* secretion data which suggests that a decrease in biosynthesis is more significant to the reduced VWF levels than any defect in secretion. It is interesting to speculate how this mutation causes a reduction in VWF levels. It is possible that despite forming WPB, this process may be less efficient in this mutant. Some VWF misfolding would be expected as a result of the missing exons and disruption to intrachain disulphide bond formation and this may compromise packaging and tubulation within the WPB. It is unclear if the WPB formed by rVWF33-34del are morphologically different to WT, despite our observation that numerous small round WPB were found in this mutant. Interestingly the rounded pseudo-WPB produced in this mutant often appeared to have a more uniform fluorescent signal, perhaps suggesting disordered VWF tubule formation and therefore a reduction in density. Potentially the exon 33-34 deletion in the D4 assembly may compromise tubule rigidity which is an important factor in WPB length (Berriman et al., 2009). Mutations in the CK domain have been proposed to disrupt the matrix between tubules and thus disrupt tubule rigidity in a similar manner (Wang et al., 2011). Both the D4 and CK domains are part of the dimeric bouquet, however it is unclear whether this region has any structural significance in terms of tubule formation and stability. Further analysis and quantification may reveal quantitative and/or morphological defects that could impact on VWF exocytosis and string formation *in vivo*. Other possibilities may include defective WPB exocytosis and impaired string formation *in vivo*.

In summary this study has demonstrated that pathogenic CNV in type 1 VWD cause the quantitative defect through ER retention and impaired secretion of VWF. These biosynthesis and secretion defects also give rise to defective WPB formation in two of the deletion mutants described here. For rVWF3del, this retention and severe reduction in secretion appears to be due to loss of the signal peptide cleavage site, representing a novel disease mechanism in type 1 VWD. While equally severe, the rVWF32-34del appears to impart its pathogenic effects in a different manner, with the apparent formation of larger misfolded VWF aggregations in the ER. The third CNV deletion rVWF33-34del showed a less severe secretion defect as supported by the formation of apparently normal pseudo-WPB. The defect in levels for this mutant appears also to be caused in

part by reduced VWF biosynthesis, however defective WPB exocytosis and aberrant string formation *in vivo* cannot be ruled out. Together these results suggest that a combination of severe reductions in secretion and biosynthesis are involved in the reduced VWF levels observed in these patients.

6 Final discussion and future work

6.1 Summary

The genetic aetiology of type 1 VWD is complex. Previous studies aimed at determining the molecular genetic and phenotypic features of type 1 VWD failed to identify causative mutations in ~35% of IC (Cumming et al., 2006, Goodeve et al., 2007, James et al., 2007). The methods of analysis used in these studies were unable to detect large heterozygous deletions due to the masking effect of the normal allele. Subsequent studies that initially identified homozygous deletions in type 3 VWD led to the realisation that large in-frame heterozygous deletions contributed to the type 1 VWD mutation spectrum (Sutherland et al., 2009). Follow up studies by the Sheffield Haemostasis Research Group, using MLPA to detect exonic CNV at the *VWF* locus, led to further identification of heterozygous deletions (Hampshire et al., 2010, Cartwright et al., 2013). These findings suggested that CNV consisting of either large deletions or duplications may be a significant factor in type 1 VWD and that they were also likely to be under-reported. Despite this finding, there remained a significant proportion of mutation-negative and CNV-negative patients.

In chapter 3 of this thesis, it was hypothesised that the reasons for this were two-fold: firstly, that a proportion of CNVs within *VWF* would remain undetected due to limitations in the MLPA probe design and secondly, that non-coding intronic and/or intergenic CNV in potential regulatory regions of the wider *VWF* locus would not have been detected due to the exon targeted nature of MLPA. It was further hypothesised that these two limitations may also apply to patients with type 2 and type 3 VWD for whom no prior mutation had been detected or where the identified mutation did not fully explain the phenotype. Based on these hypotheses, a custom microarray was designed containing 1927 *VWF* specific probes spanning the entire *VWF* coding region and extending across the 5' and 3' intergenic regions up to *CD9* and *ANO2* respectively. In addition, the design incorporated >40,000 ISCA backbone probes, giving whole genome coverage at ~41 Kb median probe spacing (~33 Kb in RefSeq genes).

Validation studies using known CNV samples demonstrated the efficacy of the design in the identification of deletions and duplications and highlighted a major advantage of aCGH over MLPA in terms of accurately identifying likely breakpoint regions. This accuracy of estimated breakpoint positions accelerates the

breakpoint mapping process and can usually be achieved in a single PCR from the two aCGH flanking probes.

Following the analysis of a mutation-negative cohort of 84 VWD samples, comprising 47 type 1 VWD families, 9 type 2 VWD individuals and 14 type 3 VWD individuals, CNV were detected in 22 cases (~26%). Interestingly in addition to exonic CNV, two intergenic CNV were also detected in this study. The exonic CNVs identified included primarily heterozygous deletions involving exons 4-5, 19-20, 33-34 in type 1 patients and exons 1-3, 1-5, 6 (homozygous), 19-20 and whole gene deletions in type 3 patients. All these deletions have been identified previously in type 1 VWD (Sutherland et al., 2009, Bowman et al., 2013, Cartwright et al., 2013) and type 3 VWD respectively (Mohl et al., 2008, Yadegari et al., 2011, Bowman et al., 2013), however it is noteworthy that the exon 33-34 deletions identified in the present study were previously undetected on a custom array that lacked sufficient pseudogene region coverage (Bellissimo et al., 2011). Additionally, in one type 1 patient, an apparent novel duplication of exons 23-34 was detected. Further work to confirm this result is ongoing. The only CNV identified in the small group of type 2 patients was an intergenic deletion.

These findings support those of previous studies and add to the growing body of evidence that suggests that large heterozygous exonic deletions are a significant factor in type 1 VWD. Previous estimates of prevalence were found to be around 6% in an EU, mutation-negative type 1 VWD cohort (Cartwright, 2013). In the present study, containing type 1 samples from the US and Canada, exonic CNV were detected in 15% of patients analysed and intergenic CNV were detected in 15%. This increased to ~50% for large exonic deletions in the small group of 14 type 3 patients.

The two 5' intergenic CNVs identified in this study (5'del-I and 5'del-II) and discussed in chapter 4 have not been reported previously in relation to VWD. Following their initial identification, the 1000 Genomes database reported frequencies for each of these deletions as 3.1% and 1.8% respectively. In 2015 towards the end of the present study, phase three of the 1000 Genomes reported frequencies of 3.4% and 3.8%. The aCGH analysis of type 1 VWD patients in this study demonstrated significantly higher frequencies of 6.2% and 12.3% for each

deletion. Based on this frequency information and the fact that these were the only non-coding CNV identified in this study, it was hypothesised that these two deletions contributed to the quantitative deficiency phenotype via disruption of regulatory elements involved in *VWF* expression. Therefore, screening for these deletions in a HC cohort was expected to yield a lower frequency, similar to that of the 1000 Genomes data. However, this was not the case, with frequencies of the HC being comparable to that of the type 1 VWD cohort, albeit a relatively small VWD cohort. This may suggest that these deletions are more frequent in the European population, which could demonstrate the differences in CNV known to exist between different populations (Sudmant et al., 2015a). Population specific CNV deletions that affect gene expression, have been described for other genetic loci including the *GSTM1* locus (Huang et al., 2009). In the relatively small number of HC analysed in the present study, no correlation with *VWF* levels could be demonstrated. However, at present this is insufficient evidence to suggest that these deletions do not influence *VWF* levels or the quantitative VWD phenotype in type 1 VWD patients. A screen of these intergenic deletions in a larger VWD group would give a clearer understanding of whether or not they represent a risk factor for low *VWF* levels or are more common in VWD patients. In addition, the data generated in this study is a correlation of historically recorded *VWF*:Ag levels with the presence (or absence) of the intergenic deletion allele(s). It is anticipated however, that disruption of upstream *cis*-acting DNA elements would likely act at the level of gene expression and regulation. Therefore, mRNA levels would be required for a more definitive conclusion of the effects of these deletions on gene expression. At present therefore it is unclear to what extent the two intergenic CNV influence *VWF* gene expression and whether any effect contributes to a quantitative deficiency.

Nevertheless, identification of 5' intergenic deletions and the evidence of increased frequency in this cohort compared to the 1000 Genomes data raises some interesting questions about upstream *cis*-acting regulatory elements in *VWF* expression. The *in silico* analysis carried out in chapter 4 highlights the high regulatory potential of multiple regions in the 5' intergenic region of *VWF*. These regions may represent active enhancer regions and are viable targets for future mutation screening in mutation-negative type 1 VWD patients.

Following on from aCGH based CNV screening and identification of intergenic CNV, chapter 5 aimed to characterise the pathogenic mechanism of three exonic deletion CNVs. Secretion defects have been reported previously for type 1 VWD mutations, often caused by ER retention of the mutant protein. However there remain few studies that have fully analysed VWF biosynthesis and WPB biogenesis for CNVs. There is also a common theme in the literature, for mutants to be compared to WT. This alone may not reflect a pattern of intracellular or ER retention and this appears to be primarily due to the fact that in some cases, the WT VWF biosynthesis is greater than that of the mutant. When this biosynthesis differential is taken into account, more accurate secretion ratios are observed that indicate reduced secretion and increased retention for the mutants. This was demonstrated in chapter 5, where three CNV deletion mutants (rVWF3del, rVWF32-34del and rVWF33-34del), when expressed *in vitro*, produced significantly lower amounts of total VWF. Previous studies have highlighted reduced total VWF biosynthesis of mutant VWF caused by missense mutations (Pruss et al., 2011, Legendre et al., 2013), however this mechanism does not appear to have been considered in studies that have analysed exonic deletion CNV mutants with reduced secretion defects (Bowman, 2013, Cartwright, 2013).

As described previously, ER retention and proteosomal degradation of the mutant and WT/mutant heterodimers may be one explanation, resulting in a dominant negative effect and ~75% reduction in VWF:Ag levels (Bodó et al., 2001). However, it is noted that in type 1 VWD, the reduction in VWF levels is frequently greater than this model would imply. This is highly suggestive of other mechanisms being involved in the reduction of total VWF levels. Furthermore, recent reports of the exon 4-5 deletion have indicated that despite a severe secretion defect, no increase in intracellular retention was observed (Bowman, 2013). This study went on to demonstrate that incubation of the mutant transfected cells with the proteasome inhibitor MG-132 showed no effect on the level of intracellular lysate VWF, indicating that the lack of increase in intracellular VWF in a mutant with severe secretion defects may be due to reduced biosynthesis rather than degradation via the proteasome. This finding is interesting and under explored. It is suggestive of a mechanism whereby the cell may regulate *VWF* expression at the level of transcription and translation. Such

a mechanism is supported by the intriguing observations of Starke et al who found that mRNA levels in type 1 VWD BOECs were significantly reduced compared to control and type 2 VWD BOECs (Starke et al., 2013). This theory would make sense given the higher energy costs of protein synthesis. It would be more efficient for the cell to prevent production of the protein in the first place rather than repeated cycles of synthesis and degradation. Studies aimed at quantifying global mRNA and protein copy numbers support this idea. It is thought that ~40% of the variance in protein levels can be accounted for by mRNA levels and the remainder can be largely explained by translational rates, a finding that may suggest that protein levels are primarily regulated at the ribosome, highlighting the significance of translational control (Schwanhaussner et al., 2011).

Microscopic evaluation of the impact of *VWF* mutations is an important consideration in order to fully understand the pathophysiology of the defect. The initial experiments described in chapter 5 highlight the importance of optimising and validating heterologous expression systems, where patient HUVEC or BOEC are not used. Without this preliminary work to demonstrate that WT *VWF* is able to form pseudo-WPB, the possibility of erroneous interpretation of the mutant disease mechanism may arise. It is essential that the heterologous expression system models the endothelial biology as closely as possible in order for meaningful interpretation of results. In this study, HEK293 cells were used as the heterologous expression system and WPB biosynthesis in these cells was optimised in order to mimic that of HUVEC. *VWF* storage and WPB formation was monitored via widefield fluorescent microscopy and SIM.

Using this system, the secretion data were supported by the microscopy results for each mutant. Two of the three CNV deletion mutants (rVWF3del and rVWF32-34del) that had the most severe secretion defects were also found to have defective WPB biogenesis and *VWF* storage. The rVWF33-34del mutant maintained pseudo-WPB biogenesis, reflecting the mild secretion deficit of this mutant. This report is the first time that super resolution SIM imaging has been applied to mutant *VWF* to observe defects in WPB formation. In this study, all three CNV deletion mutations produced quite distinct *VWF* storage phenotypes, suggesting that this technique could in future have some diagnostic utility to differentiate between certain subtypes of VWD.

The utility of SIM in diagnosing platelet granule disorders has recently been demonstrated, where it was used to differentiate between Hermansky-Pudlak Syndrome patients and healthy controls (Westmoreland et al., 2016). SIM analysis of other WPB contents may also reveal novel aspects of WPB packaging. Finally, live cell imaging using SIM could help to reveal the intricacies of WPB maturation and exocytosis.

6.2 Limitations of research

Some of the limitations associated with the research methods applied in this thesis have already been highlighted, but it is important to consider these limitations and how they may impact upon the interpretation of the experimental data. As previously described in chapter 4, the true effect of the 5' intergenic deletion CNVs identified in this study are unclear. As mRNA was unavailable for analysis, no conclusions can be made about the effect of these CNVs on gene expression. Also the sample size used for assessing the frequency of these deletions in patient and control groups would need to be increased. Although over 950 HC samples were screened for these deletions, the VWD cohort consisted of just 84 samples, which included only 47 type 1 VWD families. Clearly, a greater number of VWD patients should be screened to compare the prevalence of these deletions in the patient cohort to HCs.

In terms of the *in vitro* studies, the HEK293 cell line was the best available system for performing the studies described. However, this system is not without its limitations. HEK293 cells are not endothelial cells and so cannot truly model the natural cellular environment for VWF biosynthesis, storage and secretion. Using this system to study the heterozygous situation involves co-transfection with equal quantities of WT VWF DNA, again this is not a faithful reproduction of *in vivo* heterozygous expression as some cells may be transfected with different quantities of WT and mutant DNA. Also, transfection and therefore overexpression of VWF in these cells does not model the natural situation in endothelial cells. Overexpression may put an excessive load onto the transcriptional and translational machinery and this is likely to impact upon protein synthesis, packaging and secretion.

In this study only constitutive and basal secretion have been investigated. It would therefore be of interest to observe what effect the CNV deletions had on regulated secretion. This would be of particular interest for the rVWF33-34del which was able to form pseudo-WPB.

Finally, these experiments are of course subject to the usual limitations of *in vitro* work. Primarily the fact that *in vitro*, the impact of these mutations on the clearance of VWF cannot be studied.

6.3 Future work

The findings in this thesis present a number of potential avenues that are worthy of exploration in future studies. These are discussed below.

6.3.1 Mutation screening in 5' intergenic region

Based on the *in silico* analysis of the 5' intergenic region described in chapter 4, it is clear that this region contains a wealth of active regulatory sites. This could suggest that mutations in these regions that disrupt cis-acting enhancers could impact upon *VWF* expression and thus the quantitative defects associated with type 1 VWD.

Reduced *VWF* expression may also result from defects in *VWF* transcription, mRNA splicing, mRNA stability or translation. An understanding of the functional role of *VWF* non-coding sequences may shed light on hitherto unseen pathogenic variants within these regions. Next generation sequencing (NGS) approaches may enable this type of analysis in the future.

6.3.2 ER stress and Regulated Ire1a Dependent Decay (RIDD) in VWD

The concepts of ER stress and the UPR were introduced in chapter 5 in relation to the possible molecular mechanism of the reduced biosynthesis defects observed in this study. ER stress and the UPR is induced by the accumulation of misfolded protein in the ER lumen (Schröder and Kaufman, 2005). UPR signaling involves three ER stress signaling proteins: activating transcription factor 6 (ATF6) (Haze et al., 1999), inositol requiring enzyme 1 (IRE1) (Gardner and

Walter, 2011) and PKR-like endoplasmic reticulum kinase (PERK) (Harding et al., 1999). Downstream effectors of these signalling pathways can be detected and used to measure ER stress induction. PERK activation can be assayed via Western blot for phosphorylated eIF2 α and the C/EBP homologous protein transcription factor (CHOP) production. The activation of IRE1 can be detected by measuring xbp1 splicing. Performing these experiments for the exonic deletion mutants would give some indication of their effect on ER stress. Evidence of ER stress would provide a key link between ER retention of mutant VWF and reduced biosynthesis via translational repression.

In addition, as outlined in chapter 5 (section 5.7), a number of key findings support the hypothesis that low VWF levels in some type 1 mutations may be caused by reduced *VWF* mRNA levels (Starke et al., 2013). One possible explanation for reduced mRNA levels is increased *VWF* mRNA degradation via regulated IRE1 dependent decay (RIDD). The detection of xbp1 splicing would indicate the ER stress induced activity of IRE1. Analysis of the *VWF* mRNA sequence, suggests that it contains IRE1 cleavage motifs (Oikawa et al., 2010), suggesting that *VWF* mRNA may be a target for RIDD (figure 6.1). Furthermore some of these CUGCAG motifs are present in the loop of conserved stem-loop structures required for IRE1 cleavage, as predicted by CENTROIDFOLD RNA secondary structure prediction software (<http://rtools.cbrc.jp/centroidfold/>) (Sato et al., 2009) (figure 6.2). An experimental hypothesis for this work would be that deletion CNV in VWF cause ER retention of the mutant and mutant-WT heterodimers. This ER retention induces ER stress and the unfolded protein response, leading to translational repression and *VWF* mRNA degradation via RIDD. A schematic of this proposed mechanism is shown in figure 6.3.

```

4801 AGGAGCAAGG AGUUCAUGGA GGAGGUGAUU CAGCGGAUGG AUGUGGGCCA GGACAGCAUC
4861 CACGU CACGG UG CUGCAGUA CUC CUACAUG GUGACUGUGG AGUACCCUU CAGCGAGGCA
4921 CAGUCCAAAG GGGACAUCU GCAGCGGGUG CGAGAGAUC GCUACCAGGG CGGCAACAGG
4981 ACCAACACUG GGCUGGCCU GCGGUACCUC UCUGACCACA GCUUCUUGGU CAGCCAGGGU
5041 GACCGGGAGC AGGCGCCAA CCUGGUCUAC AUGGUCACCG GAAAUCCUGC CUCUGAUGAG
5101 AUCAAGAGGC UGCCUGGAGA CAUCCAGGUG GUGCCCAUUG GAGUGGGCCC UAAUGCCAAC
5161 GUGCAGGAGC UGGAGAGGAU UGGCUGGCC AAUGCCCCUA UCCUCAUCCA GGACUUUGAG
5221 ACGCUCUCCC GAGAGGCUC UGACCUGGUG CUGCAGAGGU GCUGCUCCGG AGAGGGGCUG
5281 CAGAUCCCCA CCCUCUCCC UGCACCUGAC UGCAGCCAGC CCCUGGACGU GAUCCUUCUC
5341 CUGGAUGGCU CCUCCAGUUU CCCAGCUUCU UAUUUUGAUG AAAUGAAGAG UUUCGCCAAG
5401 GCUUUCAUUU CAAAAGCCAA UAUAGGGCCU CGUCUCACUC AGGUGUCAGU G CUGCAG UAU
5461 GGAAGCAUCA CCACCAUUGA CGUGCCAUGG AACGUGGUCC CGGAGAAAGC CCAUUUGCUG
5521 AGCCUUGUGG ACGUCAUGCA GCGGGAGGA GGGCCAGCC AAAUCGGGGA UGCCUUGGGC
5581 UUUGCUGUGC GAUACUUGAC UUCAGAAAUG CAUGGUGCCA GGCCGGGAGC CUCAAAGGCG
5641 GUGGUCAUCC UGGUCACGGA CGUCUCUGUG GAUUCAGUGG AUUCAGCAGC UGAUGCCGCC
5701 AGGUCCAACA GAGUGACAGU GUUCCCUAUU GGAAUUGGAG AUCGCUACGA UGCAGCCCAG
5761 CUACGGAUCU UGGCAGGCC AGCAGGCGAC UCCAACGUGG UGAAAGCUCCA GCGAAUCGAA
5821 GACCUCUCCA CCAUGGUCAC CUUGGGCAAU UCCUUCUCC ACAACUGUG CUCUGGAUUU
5881 GUUAGGAUUU GCAUGGAUGA GGAUGGAAU GAGAAGAGG CCGGGGACGU CUGGACCUUG
5941 CCAGACCAGU GCCACACCGU GACUUGCCAG CCAGAUGGCC AGACCUUGCU GAAGAGUCAU
6001 CGGGUCAACU GUGACCGGG GCUGAGGCCU UCGUGCCUA ACAGCCAGUC CCCUGUUAAA
6061 GUGGAAGAGA CCUGUGGCUG CCGCUGGACC UGCCCCUGCG UGUGCACAGG CAGCUCCACU
6121 CGGCACAUCG UGACCUUUGA UGGGCAGAAU UUCAAGCUGA CUGGCAGCUG UUCUUAUGUC
6181 CUAUUUCAA ACAAGGAGCA GGACCUGGAG GUGAUUCUCC AU AUGGUGC CUGCAG CCU
6241 GGAGCAAGGC AGGGCUGCAU GAAAUCCAUC GAGGUGAAGC ACAGUGCCU CUCCGUCGAG
6301 CUGCACAGUG ACAUGGAGGU GACGGUGAAU GGGAGACUGG UCUCUGUCC UUACGUGGGU
6361 GGAACAUGG AAGUCAACGU UUAUGGUGCC AUCAUGCAUG AGGUCAGAU CAAUCACCUU
6421 GGUCACAUCU UCACAUUCAC UCCACAAAAC AAU GAGUUC AA CUGCAGCU CAG CCCCAAG
6481 ACUUUUGCUU CAAAGACGUA UGGUCUGUGU GGAUCUGUG AUGAGAACGG AGCCAAUGAC
-----

```

Figure 6.1 VWF mRNA sequence showing IRE1 cleavage motifs.

Analysis of the VWF mRNA sequence revealed the presence of multiple IRE1 cleavage consensus sequences (highlighted purple). At least three of these are able to form within the loop of a conserved stem-loop structure due to the presence of self-complementary flanking sequences (green).

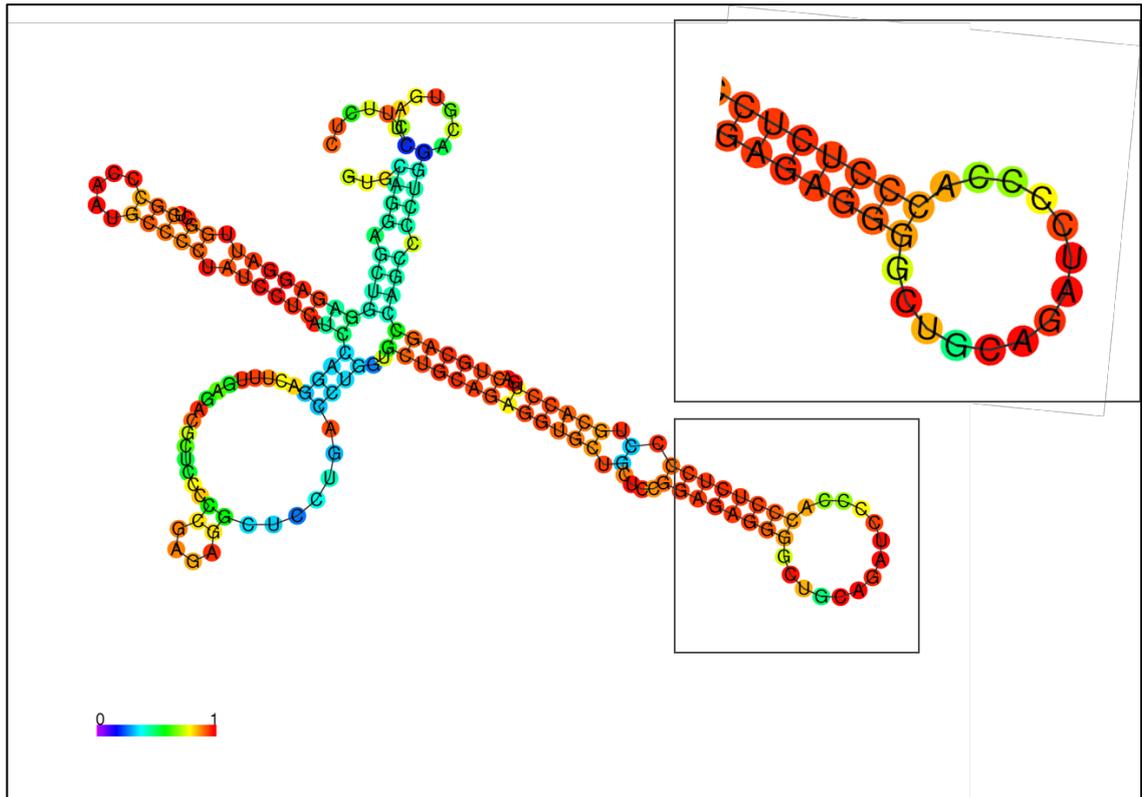


Figure 6.2 VWF mRNA secondary structure and stem loop formation.

Analysis of VWF mRNA secondary structure formation as predicted by CENTROIDFOLD software. The grey box and enlargement show the stem-loop formation with the CUGCAG cleavage site present within the loop.

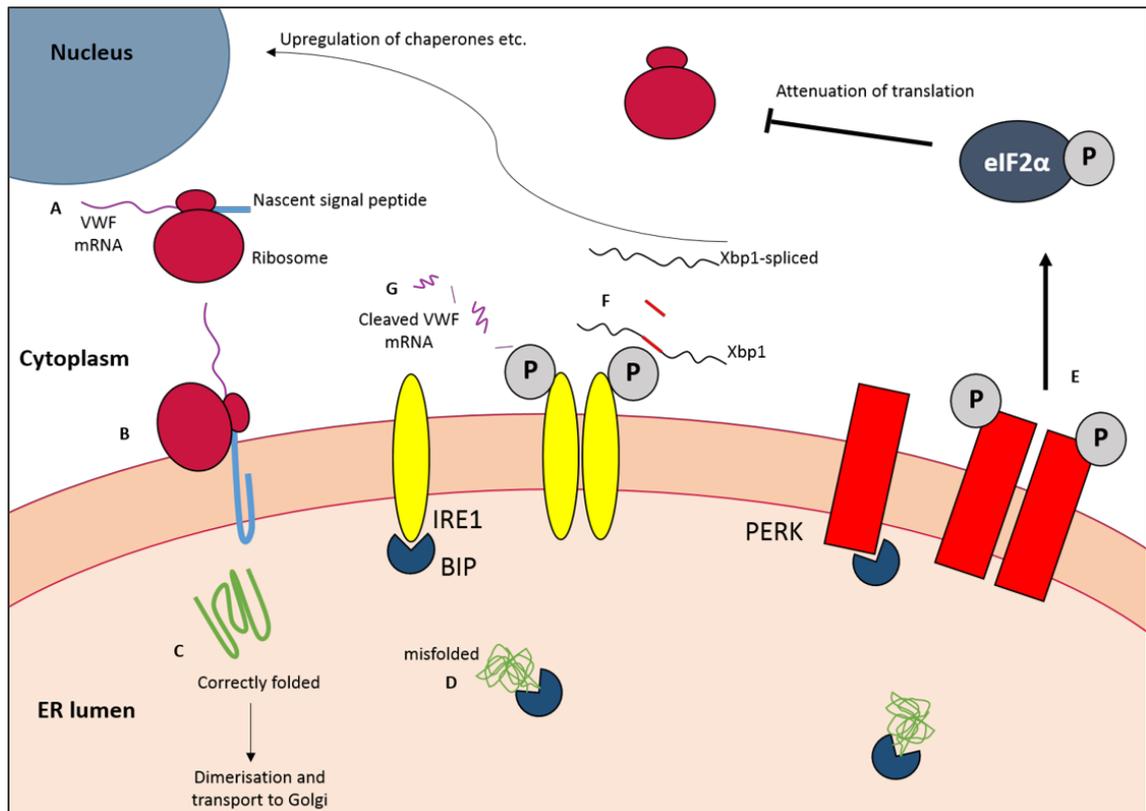


Figure 6.3 Overview of proposed ER stress effect on VWF biosynthesis.

A. VWF mRNA undergoes translation and the nascent signal peptide targets the complex to the ER. **B.** The signal peptide anchors the mRNA ribosome complex to the ER membrane where co-translational translocation into the ER lumen takes place. **C.** Under normal circumstances, the VWF folds normally, dimerises and is transported to the Golgi. **D.** Pathogenic CNV may cause misfolding of VWF and prolonged association with chaperones such as BiP. **E.** This activates the ER stress sensor PERK, which phosphorylates eIF2 α , thus shutting down translation. **F.** Accumulation of misfolded VWF may also activate IRE1 which cleaves Xbp1, producing the spliced Xbp1 mRNA which translocates to the nucleus where it encodes a transcription factor involved in upregulating ER stress related genes such as protein folding chaperones. **G.** A secondary activity of activated IRE1 is the cleavage of ER targeted mRNAs such as VWF. The ribonuclease domain of IRE1 therefore cleaves VWF mRNA during prolonged ER stress, leading to reduced VWF mRNA levels and reduced VWF biosynthesis.

6.3.3 ER stress and autophagy in VWD

Autophagy is the process of self-degradation, whereby proteins, organelles and portions of cytosol are sequestered via the formation of the autophagosome, a double membrane structure, and then delivered to the lysosome where they are degraded (He and Klionsky, 2009). ER stress can trigger autophagy (Yorimitsu et al., 2006, Hoyer-Hansen and Jaattela, 2007). ER stress and autophagy therefore appear to be linked processes as disruption of either the UPR or autophagy promotes a compensatory response from the other (Adolph et al., 2013). This is exemplified by the need for secretory cells such as pancreatic acinar cells with high protein synthesis rates, to maintain a basal level of autophagy in order to reduce ER stress and maintain protein synthesis capacity (Antonucci et al., 2015). Autophagy then, may be considered an essential homeostatic process with a vital role in secretory cells with high protein synthesis rates and thus a high protein folding load. Indeed, stimulation of autophagy using Rapamycin has been shown to alleviate ER stress in pancreatic β -cells expressing mutant pro-insulin and also in a neonatal model of hypoxia ischemia (HI) induced brain injury (Bachar-Wikstrom et al., 2013, Carloni et al., 2014). Autophagy has been shown to be down regulated due to significantly reduced Atg7 expression in a mouse model of obesity, leading to elevated ER stress and insulin resistance. Adenoviral restoration of Atg7 expression restored insulin receptor signaling and reduced obesity-induced ER stress (Yang et al., 2011). These studies demonstrate that pharmacological and genetic upregulation or restoration of autophagy, reduce ER stress and restore secretory cell homeostasis.

Together with evidence that inhibition of autophagy causes reduced VWF secretion and reduced WPB size (Torisu et al., 2013) and the identification of a large heterozygous *ATG7* deletion in a type 1 VWD patient in the present study, these data suggest that stimulation of autophagy in type 1 VWD may be a promising therapeutic approach to increase VWF levels. Further experiments should explore the role of autophagy in VWF secretion and whether it plays some role in quality control in endothelial cells in order to reduce ER stress caused by mutant or random misfolding of VWF in the ER. Although the study by Torisu et al showed reduced VWF secretion following disruption of autophagy, they did not consider what affect autophagy knock down may have on ER stress. It would be

interesting to investigate whether autophagy is necessary for reducing endothelial ER stress and conversely whether inducing autophagy with Rapamycin is able to rescue the biosynthesis and secretion defects observed in the CNV deletion mutants described in this thesis.

6.4 Concluding remarks

The molecular genetic basis of type 1 VWD is complex and there remain a number of unanswered questions regarding the full mutation spectrum and disease mechanisms. The results presented in this thesis have highlighted the utility of aCGH in the identification of novel CNV at the *VWF* locus and across the whole genome. This approach further highlighted the fact that CNV are a significant factor in the mutation spectrum of VWD and that intergenic CNV do occur at the *VWF* locus. The identification of two upstream CNV is interesting and raises important questions about the wider regulatory landscape of *VWF*.

Other data presented in this thesis attempted to unravel the molecular pathogenic mechanisms of exonic deletion CNV that cause quantitative deficiencies in VWF. The report of a deletion CNV that causes loss of the ER signal peptide cleavage site and ER retention as reported for the exon 3 deletion, highlights a novel disease mechanism in type 1 VWD. Not only did these CNV deletion variants cause reduced secretion, they also resulted in significant and profound impairment of VWF biosynthesis and WPB biogenesis as shown via SIM imaging. This is the first comparison between mutant and WT VWF storage via super-resolution microscopy and demonstrates the potential of this technique for furthering our understanding of WPB biogenesis, structure and function. These data raise important questions relating to the cause of the reduced VWF biosynthesis. Future investigations into these questions may reveal novel aspects of VWF biology in relation to health and disease.

7 References

- Adolph, T. E., Tomczak, M. F., Niederreiter, L., et al. 2013. Paneth cells as a site of origin for intestinal inflammation. *Nature*, 503, 272-276.
- Ahmad, F., Budde, U., Jan, R., et al. 2013. Phenotypic and molecular characterisation of type 3 von Willebrand disease in a cohort of Indian patients. *Thrombosis and Haemostasis*, 109, 652-660.
- Allen, S., Abuzenadah, A. M., Hinks, J., et al. 2000. A novel von Willebrand disease-causing mutation (Arg273Trp) in the von Willebrand factor propeptide that results in defective multimerization and secretion. *Blood*, 96, 560-568.
- Allen, S., Goodeve, A. C., Peake, I. R. Daly, M. E. 2001. Endoplasmic Reticulum Retention and Prolonged Association of a von Willebrand's Disease-Causing von Willebrand Factor Variant with ERp57 and Calnexin. *Biochemical and Biophysical Research Communications*, 280, 448-453.
- Allfrey, V. G., Faulkner, R. Mirsky, A. E. 1964. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences of the United States of America*, 51, 786-794.
- Antonucci, L., Fagman, J. B., Kim, J. Y., et al. 2015. Basal autophagy maintains pancreatic acinar cell homeostasis and protein synthesis and prevents ER stress. *Proceedings of the National Academy of Sciences*, 112, E6166-E6174.
- Askree, S., Chin, E., Bean, L., Coffee, B., Tanner, A. Hegde, M. 2013. Detection limit of intragenic deletions with targeted array comparative genomic hybridization. *BMC Genetics*, 14, 116.
- Assouline, Z., Kerbiriou-Nabias, D. M., Piétu, G., Thomas, N., Bahnak, B. R. Meyer, D. 1988. The human gene for von willebrand factor. Identification of repetitive ALU sequences 5' to the transcription initiation site. *Biochemical and Biophysical Research Communications*, 153, 1159-1166.

- Bachar-Wikstrom, E., Wikstrom, J. D., Ariav, Y., et al. 2013. Stimulation of Autophagy Improves Endoplasmic Reticulum Stress–Induced Diabetes. *Diabetes*, 62, 1227-1237.
- Badirou, I., Kurdi, M., Rayes, J., et al. 2010. von Willebrand factor clearance does not involve proteolysis by ADAMTS-13. *Journal of Thrombosis and Haemostasis*, 8, 2338-2340.
- Banerji, J., Olson, L. Schaffner, W. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*, 33, 729-740.
- Banerji, J., Rusconi, S. Schaffner, W. 1981. Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27, 299-308.
- Baroncini, L., Cozzi, G., Canciani, M. T., et al. 2003. Molecular defects in type 3 von Willebrand disease: updated results from 40 multiethnic patients. *Blood Cells, Molecules, and Diseases*, 30, 264-270.
- Barski, A., Cuddapah, S., Cui, K., et al. 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, 129, 823-837.
- Belfield, E. J., Brown, C., Gan, X., et al. 2014. Microarray-based ultra-high resolution discovery of genomic deletion mutations. *BMC Genomics*, 15, 1-13.
- Bell, A. C., West, A. G. Felsenfeld, G. 1999. The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. *Cell*, 98, 387-396.
- Bellissimo, D. B., Christopherson, P. A., Flood, V. H., et al. 2012. VWF mutations and new sequence variations identified in healthy controls are more frequent in the African-American population. *Blood*, 119, 2135-2140.
- Bellissimo, D. B., Ephrem, C., Hegde, M., Christopherson, P. A., Haberichter, S. L. Montgomery, R. R. Detection of VWF Copy Number Variations by Microarray Analysis in Von Willebrand Disease Type 3 Index Cases. 53rd ASH Annual Meeting and Exposition, 2011.

- Berber, E., James, P. D., Hough, C. Lillicrap, D. 2009. An assessment of the pathogenic significance of the R924Q von Willebrand factor substitution. *Journal of Thrombosis and Haemostasis*, 7, 1672-1679.
- Bernardi, F., Patracchini, P., Gemmati, D., et al. 1993. In-frame deletion of von Willebrand factor A domains in a dominant type of von Willebrand disease. *Human Molecular Genetics*, 2, 545-548.
- Berriman, J. A., Li, S., Hewlett, L. J., et al. 2009. Structural organization of Weibel-Palade bodies revealed by cryo-EM of vitrified endothelial cells. *Proceedings of the National Academy of Sciences*, 106, 17407-17412.
- Birch, K. A., Pober, J. S., Zavoico, G. B., Means, A. R. Ewenstein, B. M. 1992. Calcium/calmodulin transduces thrombin-stimulated secretion: studies in intact and minimally permeabilized human umbilical vein endothelial cells. *The Journal of Cell Biology*, 118, 1501-1510.
- Bodó, I., Katsumi, A., Tuley, E. A., Eikenboom, J. C. J., Dong, Z. Sadler, J. E. 2001. Type 1 von Willebrand disease mutation Cys1149Arg causes intracellular retention and degradation of heterodimers: a possible general mechanism for dominant mutations of oligomeric proteins. *Blood*, 98, 2973-2979.
- Boettger, L. M., Salem, R. M., Handsaker, R. E., et al. 2016. Recurring exon deletions in the HP (haptoglobin) gene contribute to lower blood cholesterol levels. *Nat Genet*, 48, 359-366.
- Boisseau, P., Ternisien, C., Caron, C., et al. 2013. Incidence of large *VWF* gene deletions and duplications in the French cohort of 1182 patients with von Willebrand disease (VWD). *Journal of Thrombosis and Haemostasis*, 11, 116-118.
- Bonthron, D. Orkin, S. H. 1988. The human von Willebrand factor gene. *European Journal of Biochemistry*, 171, 51-57.

- Bonthron, D. T., Handin, R. I., Kaufman, R. J., et al. 1986. Structure of pre-pro-von Willebrand factor and its expression in heterologous cells. *Nature*, 324, 270-273.
- Boomsma, D. I., Wijmenga, C., Slagboom, E. P., et al. 2014. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet*, 22, 221-227.
- Bowen, D. J. 2003. An influence of ABO blood group on the rate of proteolysis of von Willebrand factor by ADAMTS13. *Journal of Thrombosis and Haemostasis*, 1, 33-40.
- Bowman, M. 2013. *Investigating the Genetic Basis of Type 3 von Willebrand Disease (VWD)*. PhD, Queen's University, Kingston, Ontario, Canada.
- Bowman, M., Hopman, W. M., Rapson, D., Lillicrap, D. James, P. 2010. The prevalence of symptomatic von Willebrand disease in primary care practice. *Journal of Thrombosis and Haemostasis*, 8, 213-216.
- Bowman, M., Tuttle, A., Notley, C., et al. 2013. The genetics of Canadian type 3 von Willebrand disease: further evidence for co-dominant inheritance of mutant alleles. *Journal of Thrombosis and Haemostasis*, 11, 512-520.
- Boyer, L. A., Plath, K., Zeitlinger, J., et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, 441, 349-353.
- Boyle, A. P., Davis, S., Shulha, H. P., et al. 2009. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell*, 132, 311-322.
- Cabrera, N., Casaña, P., Cid, A. R., et al. 2011. First application of MLPA method in severe von Willebrand disease. Confirmation of a new large VWF gene deletion and identification of heterozygous carriers. *British Journal of Haematology*, 152, 240-242.

- Calton, M., Zeng, H., Urano, F., et al. 2002. IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature*, 415, 92-96.
- Carew, J., Browning, P. Lynch, D. 1990. Sulfation of von Willebrand factor. *Blood*, 76, 2530-2539.
- Carlioni, S., Albertini, M. C., Galluzzi, L., Buonocore, G., Proietti, F. Balduini, W. 2014. Increased autophagy reduces endoplasmic reticulum stress after neonatal hypoxia–ischemia: Role of protein synthesis and autophagic pathways. *Experimental Neurology*, 255, 103-112.
- Cartwright, A. 2013. *Understanding the pathogenesis of copy number variant formation in the von Willebrand factor gene*. PhD, University of Sheffield.
- Cartwright, A., Hampshire, D. J., Bloomer, L., et al. 2013. Characterisation of large in-frame deletions contributing to type 1 VWD pathogenesis in the MCMDM-1VWD study. *The XXIV Congress Of The International Society On Thrombosis And Haemostasis*.
- Carvalho, C. M. B. Lupski, J. R. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet*, 17, 224-238.
- Casari, C., Pinotti, M., Lancellotti, S., et al. 2010. The dominant-negative von Willebrand factor gene deletion p.P1127_C1948delinsR: molecular mechanism and modulation. *Blood*, 116, 5371-5376.
- Casonato, A., Pontara, E., Sartorello, F., et al. 2002. Reduced von Willebrand factor survival in type Vicenza von Willebrand disease. *Blood*, 99, 180-184.
- Castaman, G., Eikenboom, J. C. J., Bertina, R. M. Rodeghiero, F. 1999. Inconsistency of Association between Type 1 von Willebrand Disease Phenotype and Genotype in Families Identified in an Epidemiological Investigation. *Thrombosis and Haemostasis*, 82, 1065-1070.

- Castaman, G., Giacomelli, S. H., Jacobi, P. M., et al. 2012. Reduced von Willebrand factor secretion is associated with loss of Weibel–Palade body formation. *Journal of Thrombosis and Haemostasis*, 10, 951-958.
- Castaman, G., Lethagen, S., Federici, A. B., et al. 2008. Response to desmopressin is influenced by the genotype and phenotype in type 1 von Willebrand disease (VWD): results from the European Study MCMDM-1VWD. *Blood*, 111, 3531-3539.
- Castaman, G., Rodeghiero, F., Tassetto, A., et al. 2006. Hemorrhagic symptoms and bleeding risk in obligatory carriers of type 3 von Willebrand disease: an international, multicenter study. *Journal of Thrombosis and Haemostasis*, 4, 2164-2169.
- Chung, J. H., Whiteley, M. Felsenfeld, G. 1993. A 5' element of the chicken β -globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell*, 74, 505-514.
- Collins, C. J., Underdahl, J. P., Levene, R. B., et al. 1987. Molecular cloning of the human gene for von Willebrand factor and identification of the transcription initiation site. *Proceedings of the National Academy of Sciences of the United States of America*, 84, 4393-4397.
- Conrad, D. F., Pinto, D., Redon, R., et al. 2010. Origins and functional impact of copy number variation in the human genome. *Nature*, 464, 704-712.
- Consortium, E. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57-74.
- Cooper, G. M., Coe, B. P., Girirajan, S., et al. 2011. A copy number variation morbidity map of developmental delay. *Nat Genet*, 43, 838-846.
- Cournac, A., Koszul, R. Mozziconacci, J. 2016. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Research*, 44, 245-255.

- Creyghton, M. P., Cheng, A. W., Welstead, G. G., et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, 107, 21931-21936.
- Cui, J., Chen, W., Sun, J., et al. 2015. Competitive Inhibition of the Endoplasmic Reticulum Signal Peptidase by Non-cleavable Mutant Preprotein Cargos. *Journal of Biological Chemistry*, 290, 28131-28140.
- Cumming, A., Grundy, P., Keeney, S., et al. 2006. An investigation of the von Willebrand factor genotype in UK patients diagnosed to have type 1 von Willebrand disease. *Thrombosis and Haemostasis*, 96, 630-641.
- Daidone, V., Saga, G., Barbon, G., et al. 2015. The p.R1819_C1948delinsS mutation makes von Willebrand factor ADAMTS13-resistant and reduces its collagen-binding capacity. *British Journal of Haematology*, 170, 564-573.
- Daly, M., Bruce, D., Perry, D. J., et al. 1990. Antithrombin Dublin (-3 Val → Glu): an N-terminal variant which has an aberrant signal peptidase cleavage site. *FEBS Letters*, 273, 87-90.
- Datta, R., Waheed, A., Shah, G. N. Sly, W. S. 2007. Signal sequence mutation in autosomal dominant form of hypoparathyroidism induces apoptosis that is corrected by a chemical chaperone. *Proceedings of the National Academy of Sciences*, 104, 19989-19994.
- Davidson, C. J., Tuddenham, E. G. Mcvey, J. H. 2003. 450 million years of hemostasis. *Journal of Thrombosis and Haemostasis*, 1, 1487-1494.
- Davie, E. W. Ratnoff, O. D. 1964. Waterfall Sequence for Intrinsic Blood Clotting. *Science*, 145, 1310-1312.
- Dekker, J., Rippe, K., Dekker, M. Kleckner, N. 2002. Capturing Chromosome Conformation. *Science*, 295, 1306-1311.
- Desch, K. C., Ozel, A. B., Siemieniak, D., et al. 2013. Linkage analysis identifies a locus for plasma von Willebrand factor undetected by genome-wide

association. *Proceedings of the National Academy of Sciences*, 110, 588-593.

Dixon, J. R., Selvaraj, S., Yue, F., et al. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485, 376-380.

Eikenboom, J., Hilbert, L., Ribba, A. S., et al. 2009. Expression of 14 von Willebrand factor mutations identified in patients with type 1 von Willebrand disease from the MCMDM-1VWD study. *Journal of Thrombosis and Haemostasis*, 7, 1304-1312.

Eikenboom, J., Matsushita, T., Reitsma, P., et al. 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., Van Marion, V., Putter, H., et al. 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. *Journal of Thrombosis and Haemostasis*, 4, 774-782.

Eikenboom, J. C. J. 2001. Congenital von Willebrand disease type 3: clinical manifestations, pathophysiology and molecular biology. *Best Practice & Research Clinical Haematology*, 14, 365-379.

Ernst, J. Kellis, M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotech*, 28, 817-825.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473, 43-49.

Ewenstein, B., Inbal, A., Pober, J. Handin, R. 1990. Molecular studies of von Willebrand disease: reduced von Willebrand factor biosynthesis, storage, and release in endothelial cells derived from patients with type I von Willebrand disease [published erratum appears in *Blood* 1990 Nov 1;76(9):1901]. *Blood*, 75, 1466-1472.

- Federici, A. B., Bader, R., Pagani, S., Colibretti, M. L., De Marco, L. Mannucci, P. M. 1989. Binding of von Willebrand factor to glycoproteins Ib and IIb/IIIa complex: affinity is related to multimeric size. *British Journal of Haematology*, 73, 93-99.
- Ferraro, F., Kriston-Vizi, J., Metcalf, Daniel j., et al. 2014. A Two-Tier Golgi-Based Control of Organelle Size Underpins the Functional Plasticity of Endothelial Cells. *Developmental Cell*, 29, 292-304.
- Ferreira, V., Assouline, Z., Schwachtgen, J. L., Bahnak, B. R., Meyer, D. Kerbiriou-Nabias, D. 1993. The role of the 5'-flanking region in the cell-specific transcription of the human von Willebrand factor gene. *Biochemical Journal*, 293, 641-648.
- Foster, P. A., Fulcher, C. A., Marti, T., Titani, K. Zimmerman, T. S. 1987. A major factor VIII binding domain resides within the amino-terminal 272 amino acid residues of von Willebrand factor. *Journal of Biological Chemistry*, 262, 8443-6.
- Furlan, M., Robles, R. Lamie, 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.
- Gallinaro, L., Cattini, M. G., Sztukowska, M., et al. 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.
- Gardner, B. M. Walter, P. 2011. Unfolded Proteins Are Ire1-Activating Ligands That Directly Induce the Unfolded Protein Response. *Science*, 333, 1891-1894.
- Geysen, H., Tainer, J., Rodda, S., et al. 1987. Chemistry of antibody binding to a protein. *Science*, 235, 1184-1190.
- Giblin, J. P., Hewlett, L. J. Hannah, M. J. 2008. *Basal secretion of von Willebrand factor from human endothelial cells.*

- Gill, J., Endres-Brooks, J., Bauer, P., Marks, W. J. Montgomery, R. 1987. The effect of ABO blood group on the diagnosis of von Willebrand disease. *Blood*, 69, 1691-1695.
- Ginsburg, D., Handin, R., Bonthron, D., et al. 1985. Human von Willebrand factor (vWF): isolation of complementary DNA (cDNA) clones and chromosomal localization. *Science*, 228, 1401-1406.
- Gonzalez, E., Kulkarni, H., Bolivar, H., et al. 2005. The Influence of CCL3L1 Gene-Containing Segmental Duplications on HIV-1/AIDS Susceptibility. *Science*, 307, 1434-1440.
- Goodeve, A., Eikenboom, J., Castaman, G., et al. 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 Reviews*, 24, 123-134.
- Groeneveld, D. J., Van Bekkum, T., Cheung, K. L., et al. 2015. No evidence for a direct effect of von Willebrand factor's ABH blood group antigens on von Willebrand factor clearance. *Journal of Thrombosis and Haemostasis*, 13, 592-600.
- Groeneveld, D. J., Wang, J.-W., Mourik, M. J., et al. 2014. Storage and secretion of naturally occurring von Willebrand factor A domain variants. *British Journal of Haematology*, 167, 529-540.
- Gross, D. S. Garrard, W. T. 1988. Nuclease Hypersensitive Sites in Chromatin. *Annual Review of Biochemistry*, 57, 159-197.
- Gustafsson, M. G. L., Shao, L., Carlton, P. M., et al. 2008. Three-Dimensional Resolution Doubling in Wide-Field Fluorescence Microscopy by Structured Illumination. *Biophysical Journal*, 94, 4957-4970.

- Haberichter, S. L., Castaman, G., Budde, U., et al. 2008. Identification of type 1 von Willebrand disease patients with reduced von Willebrand factor survival by assay of the VWF propeptide in the European study: Molecular and Clinical Markers for the Diagnosis and Management of Type 1 VWD (MCMDM-1VWD). *Blood*, 111, 4979-4985.
- Hamilton, K. K. Sims, P. J. 1987. Changes in cytosolic Ca²⁺ associated with von Willebrand factor release in human endothelial cells exposed to histamine. Study of microcarrier cell monolayers using the fluorescent probe indo-1. *The Journal of Clinical Investigation*, 79, 600-608.
- Hampshire, D. J., Abuzenadah, A. M., Cartwright, A., et al. 2013. Identification and characterisation of mutations associated with von Willebrand disease in a Turkish patient cohort. *Thrombosis and Haemostasis*, 110, 264-274.
- Hampshire, D. J., Bloomer, L. D., Al-Buhairan, A. M., et al. 2010. Investigation of the Role of Copy Number Variation In the Pathogenesis of Type 1 Von Willebrand Disease. *American Society of Haematology Annual Meeting and Exposition*.
- Hampshire, D. J., Winship, P. R., Lillicrap, D., et al. Use of in silico and in vitro analysis to identify VWF 5' regulatory elements. The XXII Congress of the International Society of Thrombosis and Haemostasis, 2009 Boston, MA, USA. i-vii, 1-1268.
- Harding, H. P., Zhang, Y. Ron, D. 1999. Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature*, 397, 271-274.
- Hassenpflug, W. A., Budde, U., Obser, T., et al. 2006. Impact of mutations in the von Willebrand factor A2 domain on ADAMTS13-dependent proteolysis. *Blood*, 107, 2339-2345.
- Hawke, L., Bowman, M. L., Poon, M.-C., Scully, M.-F., Rivard, G.-E. James, P. D. 2016. Characterization of aberrant splicing of von Willebrand factor in

von Willebrand disease: an underrecognized mechanism. *Blood*, 128, 584-593.

Hayes, J. L., Tzika, A., Thygesen, H., et al. 2013. Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. *Genomics*, 102, 174-181.

Haze, K., Yoshida, H., Yanagi, H., Yura, T. Mori, K. 1999. Mammalian Transcription Factor ATF6 Is Synthesized as a Transmembrane Protein and Activated by Proteolysis in Response to Endoplasmic Reticulum Stress. *Molecular Biology of the Cell*, 10, 3787-3799.

Hazlewood, R. J., Roos, B. R., Solivan-Timpe, F., et al. 2015. Heterozygous Triplication of Upstream Regulatory Sequences Leads to Dysregulation of Matrix Metalloproteinase 19 in Patients with Cavitory Optic Disc Anomaly. *Human Mutation*, 36, 369-378.

He, C. Klionsky, D. J. 2009. Regulation Mechanisms and Signaling Pathways of Autophagy. *Annual Review of Genetics*, 43, 67-93.

Hegde, M. R., Chin, E. L. H., Mulle, J. G., Okou, D. T., Warren, S. T. Zwick, M. E. 2008. Microarray-based mutation detection in the dystrophin gene. *Human Mutation*, 29, 1091-1099.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459, 108-112.

Heintzman, N. D., Stuart, R. K., Hon, G., et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, 39, 311-318.

Hickson, N., Hampshire, D., Castaman, G., et al. 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. *Journal of Thrombosis and Haemostasis*, 9, 603-605.

- Hickson, N., Hampshire, D., Winship, P., et al. 2010. von Willebrand factor variant p.Arg924Gln marks an allele associated with reduced von Willebrand factor and factor VIII levels. *Journal of Thrombosis and Haemostasis*, 8, 1986-1993.
- Hill-Eubanks, D. C., Parker, C. G. Lollar, P. 1989. Differential proteolytic activation of factor VIII-von Willebrand factor complex by thrombin. *Proceedings of the National Academy of Sciences of the United States of America*, 86, 6508-6512.
- Hindorff, L. A., Sethupathy, P., Junkins, H. A., et al. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106, 9362-9367.
- Hollien, J., Lin, J. H., Li, H., Stevens, N., Walter, P. Weissman, J. S. 2009. Regulated Ire1-dependent decay of messenger RNAs in mammalian cells. *The Journal of Cell Biology*, 186, 323-331.
- Hollien, J. Weissman, J. S. 2006. Decay of Endoplasmic Reticulum-Localized mRNAs During the Unfolded Protein Response. *Science*, 313, 104-107.
- Hong, J.-W., Hendrix, D. A. Levine, M. S. 2008. Shadow Enhancers as a Source of Evolutionary Novelty. *Science*, 321, 1314-1314.
- Hoyer-Hansen, M. Jaattela, M. 2007. Connecting endoplasmic reticulum stress to autophagy by unfolded protein response and calcium. *Cell Death Differ*, 14, 1576-1582.
- Huang, R.-H., Wang, Y., Roth, R., et al. 2008. Assembly of Weibel–Palade body-like tubules from N-terminal domains of von Willebrand factor. *Proceedings of the National Academy of Sciences*, 105, 482-487.
- Huang, R. S., Chen, P., Wisel, S., et al. 2009. Population-specific GSTM1 copy number variation. *Human Molecular Genetics*, 18, 366-372.

- Huizinga, E. G., Tsuji, S., Romijn, R. a. P., et al. 2002. Structures of Glycoprotein I β and Its Complex with von Willebrand Factor A1 Domain. *Science*, 297, 1176-1179.
- lafrate, A. J., Feuk, L., Rivera, M. N., et al. 2004. Detection of large-scale variation in the human genome. *Nat Genet*, 36, 949-951.
- Ito, M., Oiso, Y., Murase, T., et al. 1993. Possible involvement of inefficient cleavage of preprovasopressin by signal peptidase as a cause for familial central diabetes insipidus. *The Journal of Clinical Investigation*, 91, 2565-2571.
- Jahroudi, N., Ardekani, A. M. Greenberger, J. S. 1996. An NF1-like Protein Functions as a Repressor of the von Willebrand Factor Promoter. *Journal of Biological Chemistry*, 271, 21413-21421.
- Jahroudi, N. Lynch, D. C. 1994. Endothelial-cell-specific regulation of von Willebrand factor gene expression. *Molecular and Cellular Biology*, 14, 999-1008.
- James, P. D. Lillicrap, D. 2013. The molecular characterization of von Willebrand disease: good in parts. *British Journal of Haematology*, 161, 166-176.
- James, P. D., Notley, C., Hegadorn, C., et al. 2007. The mutational spectrum of type 1 von Willebrand disease: results from a Canadian cohort study. *Blood*, 109, 145-154.
- James, P. D., Paterson, A. D., Notley, C., et al. 2006. Genetic linkage and association analysis in type 1 von Willebrand disease: results from the Canadian Type 1 VWD Study. *Journal of Thrombosis and Haemostasis*, 4, 783-792.
- Johansson, A. M., Lanke, E., Säll, T., Lethagen, S. Halldén, C. 2011. A large deletion identified in a Swedish family with type 1 VWD. *Thrombosis and Haemostasis*, 105, 733-734.

- Johansson, L. F., Van Dijk, F., De Boer, E. N., et al. 2016. CoNVaDING: Single Exon Variation Detection in Targeted NGS Data. *Human Mutation*, 37, 457-464.
- Julià, A., Pinto, J. A., Gratacós, J., et al. 2015. A deletion at ADAMTS9-MAGI1 locus is associated with psoriatic arthritis risk. *Annals of the Rheumatic Diseases*, 74, 1875-1881.
- Junier, I., Dale, R. K., Hou, C., Képès, F. Dean, A. 2012. CTCF-mediated transcriptional regulation through cell type-specific chromosome organization in the β -globin locus. *Nucleic Acids Research*, 40, 7718-7727.
- Kallioniemi, A., Kallioniemi, O., Sudar, D., et al. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, 258, 818-821.
- Kamino, H., Yamazaki, Y., Saito, K., et al. 2011. Nuclear receptor CAR-regulated expression of the FAM84A gene during the development of mouse liver tumors. *International journal of oncology*, 38, 1511-1520.
- Kasatkar, P., Shetty, S. Ghosh, K. 2014. Genetic Heterogeneity in a Large Cohort of Indian Type 3 von Willebrand Disease Patients. *PLoS ONE*, 9, e92575.
- Katsumi, A., Tuley, E. A., Bodó, I. Sadler, J. E. 2000. Localization of Disulfide Bonds in the Cystine Knot Domain of Human von Willebrand Factor. *Journal of Biological Chemistry*, 275, 25585-25594.
- Kleinschmidt, A. M., Nassiri, M., Stitt, M. S., et al. 2008. Sequences in Intron 51 of the von Willebrand Factor Gene Target Promoter Activation to a Subset of Lung Endothelial Cells in Transgenic Mice. *Journal of Biological Chemistry*, 283, 2741-2750.
- Klenow, H. Henningsen, I. 1970. Selective Elimination of the Exonuclease Activity of the Deoxyribonucleic Acid Polymerase from Escherichia coli B by Limited Proteolysis. *Proceedings of the National Academy of Sciences*, 65, 168-175.

- Klopocki, E. Mundlos, S. 2011. Copy-Number Variations, Noncoding Sequences, and Human Phenotypes. *Annual Review of Genomics and Human Genetics*, 12, 53-72.
- Kobayashi, T., Masaki, T., Sugiyama, M., Atomi, Y., Furukawa, Y. Nakamura, Y. 2006. A gene encoding a family with sequence similarity 84, member A (FAM84A) enhanced migration of human colon cancer cells. *International Journal of Oncology*, 29, 341-347.
- Kornberg, R. D. 1974. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science*, 184, 868-871.
- Kornberg, R. D. Thonmas, J. O. 1974. Chromatin Structure: Oligomers of the Histones. *Science*, 184, 865-868.
- Kouzarides, T. 2007. Chromatin Modifications and Their Function. *Cell*, 128, 693-705.
- Kozutsumi, Y., Segal, M., Normington, K., Gething, M.-J. Sambrook, J. 1988. The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins. *Nature*, 332, 462-464.
- Lakkaraju, A. K. K., Mary, C., Scherrer, A., Johnson, A. E. Strub, K. 2008. SRP Keeps Polypeptides Translocation-Competent by Slowing Translation to Match Limiting ER-Targeting Sites. *Cell*, 133, 440-451.
- Legendre, P., Navarrete, A.-M., Rayes, J., et al. 2013. Mutations in the A3 domain of Von Willebrand factor inducing combined qualitative and quantitative defects in the protein. *Blood*, 121, 2135-2143.
- Lek, M., Karczewski, K. J., Minikel, E. V., et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536, 285-291.
- Lemaire, K., Moura, R. F., Granvik, M., et al. 2011. Ubiquitin Fold Modifier 1 (UFM1) and Its Target UFBP1 Protect Pancreatic Beta Cells from ER Stress-Induced Apoptosis. *PLoS ONE*, 6, e18517.

- Lenting, P. J., Westein, E., Terraube, V., et al. 2004. An Experimental Model to Study the in Vivo Survival of von Willebrand Factor: BASIC ASPECTS AND APPLICATION TO THE R1205H MUTATION. *Journal of Biological Chemistry*, 279, 12102-12109.
- Lentz, S. Sadler, J. 1993. Homocysteine inhibits von Willebrand factor processing and secretion by preventing transport from the endoplasmic reticulum. *Blood*, 81, 683-689.
- Levy, G. G., Nichols, W. C., Lian, E. C., et al. 2001. Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura. *Nature*, 413, 488-494.
- Li, Q. Stamatoyannopoulos, G. 1994. Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator. *Blood*, 84, 1399-1401.
- Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., et al. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326, 289-293.
- Lindstrand, A., Frangakis, S., Carvalho, Claudia m. B., et al. 2016. Copy-Number Variation Contributes to the Mutational Load of Bardet-Biedl Syndrome. *The American Journal of Human Genetics*, 99, 318-336.
- Lippok, S., Kolšek, K., Löf, A., et al. 2016. von Willebrand factor is dimerized by protein disulfide isomerase. *Blood*, 127, 1183-1191.
- Lopes Da Silva, M. Cutler, D. F. 2016. von Willebrand factor multimerization and the polarity of secretory pathways in endothelial cells. *Blood*, 128, 277-285.
- Lopes Da Silva, M., O'connor, M. N., Kriston-Vizi, J., et al. 2016. Type II PI4-kinases control Weibel-Palade body biogenesis and von Willebrand factor structure in human endothelial cells. *Journal of Cell Science*, 129, 2096-2105.

- Lupiáñez, Darío g., Kraft, K., Heinrich, V., et al. 2015. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, 161, 1012-1025.
- Lynch, D. C., Zimmerman, T. S., Collins, C. J., et al. 1985. Molecular cloning of cDNA for human von willebrand factor: Authentication by a new method. *Cell*, 41, 49-56.
- Macdonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L. Scherer, S. W. 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research*, 42, D986-D992.
- Macfarlane, R. G. 1964. An Enzyme Cascade in the Blood Clotting Mechanism, and its Function as a Biochemical Amplifier. *Nature*, 202, 498-499.
- Mancuso, D. J., Tuley, E. A., Castillo, R., De Bosch, N., Mannucci, P. M. Sadler, J. E. 1994. Characterization of partial gene deletions in type III von Willebrand disease with alloantibody inhibitors. *Thrombosis and haemostasis*, 72, 180-185.
- Mancuso, D. J., Tuley, E. A., Westfield, L. A., et al. 1989. Structure of the gene for human von Willebrand factor. *Journal of Biological Chemistry*, 264, 19514-27.
- Manolio, T. A., Collins, F. S., Cox, N. J., et al. 2009. Finding the missing heritability of complex diseases. *Nature*, 461, 747-753.
- Marti, T., Rosselet, S. J., Titani, K. Walsh, K. A. 1987. Identification of disulfide-bridged substructures within human von Willebrand factor. *Biochemistry*, 26, 8099-8109.
- Maston, G. A., Evans, S. K. Green, M. R. 2006. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*, 7, 29-59.
- 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. *Journal of Biological Chemistry*, 267, 8723-8731.

- Mayadas, T. N. Wagner, D. D. 1992. Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. *Proceedings of the National Academy of Sciences*, 89, 3531-3535.
- Mccarroll, S. A., Huett, A., Kuballa, P., et al. 2008. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet*, 40, 1107-1112.
- Mckinnon, T. a. J., Goode, E. C., Birdsey, G. M., et al. 2010. Specific N-linked glycosylation sites modulate synthesis and secretion of von Willebrand factor. *Blood*, 116, 640-648.
- Melo, U. S., Macedo-Souza, L. I., Figueiredo, T., et al. 2015. Overexpression of KLC2 due to a homozygous deletion in the non-coding region causes SPOAN syndrome. *Human Molecular Genetics*, 24, 6877-6885.
- Meyer, D. I., Krause, E. Dobberstein, B. 1982. Secretory protein translocation across membranes-the role of the 'docking protein'. *Nature*, 297, 647-650.
- Michaux, G., Abbitt, K. B., Collinson, L. M., Haberichter, S. L., Norman, K. E. Cutler, D. F. 2006. The Physiological Function of von Willebrand's Factor Depends on Its Tubular Storage in Endothelial Weibel-Palade Bodies. *Developmental Cell*, 10, 223-232.
- Michaux, G., Hewlett, L. J., Messenger, S. L., et al. 2003. Analysis of intracellular storage and regulated secretion of 3 von Willebrand disease-causing variants of von Willebrand factor. *Blood*, 102, 2452-2458.
- Miller, D. T., Adam, M. P., Aradhya, S., et al. 2010. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. *The American Journal of Human Genetics*, 86, 749-764.

- Mohl, A., Marschalek, R., Masszi, T., et al. 2008. An Alu-mediated novel large deletion is the most frequent cause of type 3 von Willebrand disease in Hungary. *Journal of Thrombosis and Haemostasis*, 6, 1729-1735.
- Monroe, D. M. Hoffman, M. 2006. What Does It Take to Make the Perfect Clot? *Arteriosclerosis, Thrombosis, and Vascular Biology*, 26, 41-48.
- Montgomery, R. R. 2006. When it comes to von Willebrand disease, does $1 + 1 = 3$? *Journal of Thrombosis and Haemostasis*, 4, 2162-2163.
- Mourik, M. J., Faas, F. G. A., Zimmermann, H., Eikenboom, J. Koster, A. J. 2015a. Towards the imaging of Weibel–Palade body biogenesis by serial block face-scanning electron microscopy. *Journal of Microscopy*, n/a-n/a.
- Mourik, M. J., Faas, F. G. A., Zimmermann, H., Voorberg, J., Koster, A. J. Eikenboom, J. 2015b. Content delivery to newly forming Weibel-Palade bodies is facilitated by multiple connections with the Golgi apparatus. *Blood*, 125, 3509-3516.
- Mourik, M. J., Valentijn, J. A., Voorberg, J., Koster, A. J., Valentijn, K. M. Eikenboom, J. 2013. von Willebrand factor remodeling during exocytosis from vascular endothelial cells. *Journal of Thrombosis and Haemostasis*, 11, 2009-2019.
- Nassiri, M., Liu, J., Kulak, S., et al. 2010. Repressors NFI and NFY Participate in Organ-Specific Regulation of von Willebrand Factor Promoter Activity in Transgenic Mice. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 30, 1423-1429.
- Ngo, K. Y., Glotz, V. T., Koziol, J. A., et al. 1988. Homozygous and heterozygous deletions of the von Willebrand factor gene in patients and carriers of severe von Willebrand disease. *Proceedings of the National Academy of Sciences of the United States of America*, 85, 2753-2757.
- Nichols, W. L., Hultin, M. B., James, A. H., et al. 2008. von Willebrand disease (VWD): evidence-based diagnosis and management guidelines, the

National Heart, Lung, and Blood Institute (NHLBI) Expert Panel report (USA)¹. *Haemophilia*, 14, 171-232.

Nik-Zainal, S., Cotter, P. E., Willatt, L. R., Abbott, K. O'Brien, E. W. 2011. Ring chromosome 12 with inverted microduplication of 12p13.3 involving the Von Willebrand Factor gene associated with cryptogenic stroke in a young adult male. *European Journal of Medical Genetics*, 54, 97-101.

Nilsson, I. M., Blombäck, M., Jorpes, E., Blombäck, B. Johansson, S.-A. 1957. v. Willebrand's Disease and its Correction with Human Plasma Fraction 1-0. *Acta Medica Scandinavica*, 159, 179-188.

Nishihara, H., Smit, A. F. A. Okada, N. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Research*, 16, 864-874.

Nossent, A. Y., Robben, J. H., Deen, P. M. T., et al. 2010. Functional variation in the arginine vasopressin 2 receptor as a modifier of human plasma von Willebrand factor levels. *Journal of Thrombosis and Haemostasis*, 8, 1547-1554.

O'Brien, L. A., James, P. D., Othman, M., et al. 2003. Founder von Willebrand factor haplotype associated with type 1 von Willebrand disease. *Blood*, 102, 549-557.

Obser, T., Ledford-Kraemer, M., Oyen, F., et al. 2016. Identification and characterization of the elusive mutation causing the historical von Willebrand Disease type IIC Miami. *Journal of Thrombosis and Haemostasis*, 14, 1725-1735.

Oikawa, D., Tokuda, M., Hosoda, A. Iwawaki, T. 2010. Identification of a consensus element recognized and cleaved by IRE1 α . *Nucleic Acids Research*, 38, 6265-6273.

Othman, M., Chirinian, Y., Brown, C., et al. 2010. Functional characterization of a 13-bp deletion (c.-1522_-1510del13) in the promoter of the von

- Willebrand factor gene in type 1 von Willebrand disease. *Blood*, 116, 3645-3652.
- Oudet, P., Gross-Bellard, M. Chambon, P. 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, 4, 281-300.
- Outinen, P. A., Sood, S. K., Pfeifer, S. I., et al. 1999. Homocysteine-Induced Endoplasmic Reticulum Stress and Growth Arrest Leads to Specific Changes in Gene Expression in Human Vascular Endothelial Cells. *Blood*, 94, 959-967.
- Pang, A., Macdonald, J., Pinto, D., et al. 2010. Towards a comprehensive structural variation map of an individual human genome. *Genome Biology*, 11, R52.
- Peake, I., Liddell, M., Moodie, P., et al. 1990. Severe type III von Willebrand's disease caused by deletion of exon 42 of the von Willebrand factor gene: family studies that identify carriers of the condition and a compound heterozygous individual. *Blood*, 75, 654-661.
- Peng, Y. Jahroudi, N. 2003. The NFY Transcription Factor Inhibits von Willebrand Factor Promoter Activation in Non-endothelial Cells through Recruitment of Histone Deacetylases. *Journal of Biological Chemistry*, 278, 8385-8394.
- Perry, G. H., Dominy, N. J., Claw, K. G., et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet*, 39, 1256-1260.
- Petersen, T. N., Brunak, S., Von Heijne, G. Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth*, 8, 785-786.
- Pinkel, D. Albertson, D. G. 2005. COMPARATIVE GENOMIC HYBRIDIZATION. *Annual Review of Genomics and Human Genetics*, 6, 331-354.

- Pinkel, D., Se Graves, R., Sudar, D., et al. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*, 20.
- Pruss, C. M., Golder, M., Bryant, A., et al. 2011. Pathologic mechanisms of type 1 VWD mutations R1205H and Y1584C through in vitro and in vivo mouse models. *Blood*, 117, 4358-4366.
- Purvis, A. R. Sadler, J. E. 2004. A Covalent Oxidoreductase Intermediate in Propeptide-dependent von Willebrand Factor Multimerization. *Journal of Biological Chemistry*, 279, 49982-49988.
- Racchi, M., Watzke, H. H., High, K. A. Lively, M. O. 1993. Human coagulation factor X deficiency caused by a mutant signal peptide that blocks cleavage by signal peptidase but not targeting and translocation to the endoplasmic reticulum. *Journal of Biological Chemistry*, 268, 5735-40.
- Rastegarlar, G., Pegon, J. N., Casari, C., et al. 2012. Macrophage LRP1 contributes to the clearance of von Willebrand factor. *Blood*, 119, 2126-2134.
- Rawley, O., O'sullivan, J. M., Chion, A., et al. 2015. von Willebrand factor arginine 1205 substitution results in accelerated macrophage-dependent clearance in vivo. *Journal of Thrombosis and Haemostasis*, 13, 821-826.
- Redon, R., Ishikawa, S., Fitch, K. R., et al. 2006. Global variation in copy number in the human genome. *Nature*, 444, 444-454.
- Rodeghiero, F., Castaman, G. Dini, E. 1987. Epidemiological investigation of the prevalence of von Willebrand's disease. *Blood*, 69, 454-459.
- Romijn, R. a. P., Bouma, B., Wuyster, W., et al. 2001. Identification of the Collagen-binding Site of the von Willebrand Factor A3-domain. *Journal of Biological Chemistry*, 276, 9985-9991.

- Ruggeri, Z. M., De Marco, L., Gatti, L., Bader, R. Montgomery, R. R. 1983. Platelets have more than one binding site for von Willebrand factor. *The Journal of Clinical Investigation*, 72, 1-12.
- Rydz, N., Swystun, L. L., Notley, C., et al. 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.
- Sadler, J. E. 1998. BIOCHEMISTRY AND GENETICS OF VON WILLEBRAND FACTOR. *Annual Review of Biochemistry*, 67, 395-424.
- Sadler, J. E., Budde, U., Eikenboom, J. C. J., et al. 2006. Update on the pathophysiology and classification of von Willebrand disease: a report of the Subcommittee on von Willebrand Factor. *Journal of Thrombosis and Haemostasis*, 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. *Proceedings of the National Academy of Sciences*, 82, 6394-6398.
- Sankaran, V. G., Menne, T. F., Xu, J., et al. 2008. Human Fetal Hemoglobin Expression Is Regulated by the Developmental Stage-Specific Repressor *BCL11A*. *Science*, 322, 1839-1842.
- Sankaran, V. G., Xu, J., Byron, R., et al. 2011. A Functional Element Necessary for Fetal Hemoglobin Silencing. *New England Journal of Medicine*, 365, 807-814.
- Sato, K., Hamada, M., Asai, K. Mituyama, T. 2009. CentroidFold: a web server for RNA secondary structure prediction. *Nucleic Acids Research*, 37, W277-W280.
- Savage, B., Saldívar, E. Ruggeri, Z. M. 1996. Initiation of Platelet Adhesion by Arrest onto Fibrinogen or Translocation on von Willebrand Factor. *Cell*, 84, 289-297.

- Schindelin, J., Arganda-Carreras, I., Frise, E., et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Meth*, 9, 676-682.
- Schneppenheim, R., Brassard, J., Krey, S., et al. 1996. Defective dimerization of von Willebrand factor subunits due to a Cys-> Arg mutation in type IID von Willebrand disease. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 3581-3586.
- Schneppenheim, R., Castaman, G., Federici, A. B., et al. 2007. A common 253-kb deletion involving VWF and TMEM16B in German and Italian patients with severe von Willebrand disease type 3. *Journal of Thrombosis and Haemostasis*, 5, 722-728.
- Schneppenheim, R., Krey, S., Bergmann, F., et al. 1994. Genetic heterogeneity of severe von Willebrand disease type III in the German population. *Human Genetics*, 94, 640-652.
- Schouten, J. P., Mcelgunn, C. J., Waaijer, R., Zwijnenburg, D., Diepvens, F. Pals, G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research*, 30, e57.
- Schröder, M. Kaufman, R. J. 2005. THE MAMMALIAN UNFOLDED PROTEIN RESPONSE. *Annual Review of Biochemistry*, 74, 739-789.
- Schwachtgen, J.-L., Remacle, J. E., Janel, N., et al. 1998. *Oct-1 Is Involved in the Transcriptional Repression of the von Willebrand Factor Gene Promoter.*
- Schwachtgen, J. L., Janel, N., Barek, L., et al. 1997. Ets transcription factors bind and transactivate the core promoter of the von Willebrand factor gene. *Oncogene*, 15, 3091-3102.
- Schwanhausser, B., Busse, D., Li, N., et al. 2011. Global quantification of mammalian gene expression control. *Nature*, 473, 337-342.
- Scott, J. P., Montgomery, R. R. Retzinger, G. S. 1991. Dimeric ristocetin flocculates proteins, binds to platelets, and mediates von Willebrand

- factor-dependent agglutination of platelets. *Journal of Biological Chemistry*, 266, 8149-8155.
- Sebat, J., Lakshmi, B., Troge, J., et al. 2004. Large-Scale Copy Number Polymorphism in the Human Genome. *Science*, 305, 525-528.
- Shelton-Inloes, B. B., Chehab, F. F., Mannucci, P. M., Federici, A. B. Sadler, J. E. 1987. Gene deletions correlate with the development of alloantibodies in von Willebrand disease. *The Journal of Clinical Investigation*, 79, 1459-1465.
- Shiltagh, N., Kirkpatrick, J., Cabrita, L. D., et al. 2014. *Solution structure of the major factor VIII binding region on von Willebrand factor*.
- Smith, N. L., Chen, M.-H., Dehghan, A., et al. 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.
- Solimando, M., Baronciani, L., La Marca, S., et al. 2012. Molecular characterization, recombinant protein expression, and mRNA analysis of type 3 von Willebrand disease: Studies of an Italian cohort of 10 patients. *American Journal of Hematology*, 87, 870-874.
- Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., et al. 1997. Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances. *Genes, Chromosomes and Cancer*, 20, 399-407.
- Souto, J. C., Almasy, L., Soria, J. M., et al. 2003. Genome-wide linkage analysis of von Willebrand factor plasma levels: results from the GAIT project. *Thrombosis and Haemostasis*, 89, 468-474.
- Sporn, L. A., Chavin, S. I., Marder, V. J. Wagner, D. D. 1985. Biosynthesis of von Willebrand protein by human megakaryocytes. *The Journal of Clinical Investigation*, 76, 1102-1106.

- Sporn, L. A., Marder, V. J. Wagner, D. D. 1986. Inducible secretion of large, biologically potent von Willebrand factor multimers. *Cell*, 46, 185-190.
- Springer, T. A. 2011. Biology and physics of von Willebrand factor concatamers. *Journal of Thrombosis and Haemostasis*, 9, 130-143.
- Starke, R. D., Paschalaki, K. E., Dyer, C. E. F., et al. 2013. Cellular and molecular basis of von Willebrand disease: studies on blood outgrowth endothelial cells. *Blood*, 121, 2773-2784.
- Steele-Stallard, H. B., Le Quesne Stabej, P., Lenassi, E., et al. 2013. Screening for duplications, deletions and a common intronic mutation detects 35% of second mutations in patients with USH2A monoallelic mutations on Sanger sequencing. *Orphanet Journal of Rare Diseases*, 8, 1-11.
- Støy, J., Edghill, E. L., Flanagan, S. E., et al. 2007. Insulin gene mutations as a cause of permanent neonatal diabetes. *Proceedings of the National Academy of Sciences*, 104, 15040-15044.
- Su, M., Han, D., Boyd-Kirkup, J., Yu, X. Han, J.-Dong j. 2014. Evolution of Alu Elements toward Enhancers. *Cell Reports*, 7, 376-385.
- Sudmant, P. H., Mallick, S., Nelson, B. J., et al. 2015a. Global diversity, population stratification, and selection of human copy-number variation. *Science*, 349.
- Sudmant, P. H., Rausch, T., Gardner, E. J., et al. 2015b. An integrated map of structural variation in 2,504 human genomes. *Nature*, 526, 75-81.
- Sutherland, M. S., Cumming, A. M., Bowman, M., et al. 2009. A novel deletion mutation is recurrent in von Willebrand disease types 1 and 3. *Blood*, 114, 1091-1098.
- The 1000 Genomes Project Consortium 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56-65.
- The Genomes Project, C. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56-65.

- The Genomes Project, C. 2015. A global reference for human genetic variation. *Nature*, 526, 68-74.
- Theophilus, B., Baugh, M. L., Guillatt, A., Wilde, J., Motwani, J. Williams, M. 2013. Identification of VWF gene deletions in 9 VWD families using multiplex ligation-dependent probe amplification (MLPA). *Journal of Thrombosis and Haemostasis*, 11(Suppl.2), 311.
- Thurman, R. E., Rynes, E., Humbert, R., et al. 2012. The accessible chromatin landscape of the human genome. *Nature*, 489, 75-82.
- Tjernberg, P., Vos, H. L., Castaman, G., Bertina, R. M. Eikenboom, J. C. J. 2004. Dimerization and multimerization defects of von Willebrand factor due to mutated cysteine residues. *Journal of Thrombosis and Haemostasis*, 2, 257-265.
- Torisu, T., Torisu, K., Lee, I. H., et al. 2013. Autophagy regulates endothelial cell processing, maturation and secretion of von Willebrand factor. *Nat Med*, 19, 1281-1287.
- Travers, K. J., Patil, C. K., Wodicka, L., Lockhart, D. J., Weissman, J. S. Walter, P. 2000. Functional and Genomic Analyses Reveal an Essential Coordination between the Unfolded Protein Response and ER-Associated Degradation. *Cell*, 101, 249-258.
- Tsai, H.-M., Nagel, R. L., Hatcher, V. B., Seaton, A. C. Sussman, I. I. 1991. The high molecular weight form of endothelial cell von Willebrand factor is released by the regulated pathway. *British Journal of Haematology*, 79, 239-245.
- Tsai, Y.-C., Cooke, N. E. Liebhaber, S. A. 2014. Tissue specific CTCF occupancy and boundary function at the human growth hormone locus. *Nucleic Acids Research*, 42, 4906-4921.
- Tuddenham, E. G. D., Lane, R. S., Rotblat, F., et al. 1982. Response to infusions of polyelectrolyte fractionated human factor VIII concentrate in human

- haemophilia A and von Willebrand's disease. *British Journal of Haematology*, 52, 259-267.
- Untergasser, A., Cutcutache, I., Koressaar, T., et al. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research*, 40, e115.
- Valentijn, K. M., Van Driel, L. F., Mourik, M. J., et al. 2010. *Multigranular exocytosis of Weibel-Palade bodies in vascular endothelial cells*.
- Van Loon, J., Dehghan, A., Weihong, T., et al. 2016. Genome-wide association studies identify genetic loci for low von Willebrand factor levels. *Eur J Hum Genet*, 24, 1035-1040.
- Van Schooten, C. J., Shahbazi, S., Groot, E., et al. 2008. Macrophages contribute to the cellular uptake of von Willebrand factor and factor VIII in vivo. *Blood*, 112, 1704-1712.
- Van Schooten, C. J., Tjernberg, P., Westein, E., et al. 2005. Cysteine-mutations in von Willebrand factor associated with increased clearance. *Journal of Thrombosis and Haemostasis*, 3, 2228-2237.
- Versteeg, H. H., Heemskerk, J. W. M., Levi, M. Reitsma, P. H. 2013. New Fundamentals in Hemostasis. *Physiological Reviews*, 93, 327-358.
- Verweij, C. L., De Vries, C. J. M., Distel, B., et al. 1985. Construction of cDNA coding for human von Willebrand factor using antibody probes for colony-screening and mapping of the chromosomal gene. *Nucleic Acids Research*, 13, 4699-4717.
- Vischer, U. Wagner, D. 1994. von Willebrand factor proteolytic processing and multimerization precede the formation of Weibel-Palade bodies. *Blood*, 83, 3536-3544.
- Vogler, C., Gschwind, L., Röthlisberger, B., et al. 2010. Microarray-Based Maps of Copy-Number Variant Regions in European and Sub-Saharan Populations. *PLoS ONE*, 5, e15246.

- Von Heijne, G. 1983. Patterns of Amino Acids near Signal-Sequence Cleavage Sites. *European Journal of Biochemistry*, 133, 17-21.
- Von Heijne, G. 1985. Signal sequences. The limits of variation. *Journal of Molecular Biology*, 184, 99-105.
- Von Willebrand, E. A. 1999. Hereditary pseudohaemophilia. *Haemophilia*, 5, 223-231.
- Wagner, D. D. Marder, V. J. 1983. Biosynthesis of von Willebrand protein by human endothelial cells. Identification of a large precursor polypeptide chain. *Journal of Biological Chemistry*, 258, 2065-7.
- 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. *The Journal of Cell Biology*, 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. *The Journal of Cell Biology*, 95, 355-360.
- Walter, P. Blobel, G. 1981. Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *The Journal of Cell Biology*, 91, 557-561.
- Wang, J.-W., Bouwens, E. a. M., Pintao, M. C., et al. 2013a. Analysis of the storage and secretion of von Willebrand factor in blood outgrowth endothelial cells derived from patients with von Willebrand disease. *Blood*, 121, 2762-2772.
- Wang, J.-W., Groeneveld, D. J., Cossemans, G., et al. 2012. Biogenesis of Weibel-Palade bodies in von Willebrand's disease variants with impaired von Willebrand factor intrachain or interchain disulfide bond formation. *Haematologica*, 97, 859-866.

- Wang, J.-W., Valentijn, K. M., De Boer, H. C., et al. 2011. Intracellular Storage and Regulated Secretion of Von Willebrand Factor in Quantitative Von Willebrand Disease. *Journal of Biological Chemistry*, 286, 24180-24188.
- Wang, Q. Y., Song, J., Gibbs, R. A., Boerwinkle, E., Dong, J. F. Yu, F. L. 2013b. Characterizing polymorphisms and allelic diversity of von Willebrand factor gene in the 1000 Genomes. *Journal of Thrombosis and Haemostasis*, 11, 261-269.
- Weedon, M. N., Cebola, I., Patch, A.-M., et al. 2014. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet*, 46, 61-64.
- Weibel, E. R. Palade, G. E. 1964. NEW CYTOPLASMIC COMPONENTS IN ARTERIAL ENDOTHELIA. *The Journal of Cell Biology*, 23, 101-112.
- Weiss, H. J., Sussman, I. I. Hoyer, L. W. 1977. Stabilization of Factor VIII in Plasma by the von Willebrand Factor: STUDIES ON POSTTRANSFUSION AND DISSOCIATED FACTOR VIII AND IN PATIENTS WITH VON WILLEBRAND'S DISEASE. *The Journal of Clinical Investigation*, 60, 390-404.
- Werner, E. J., Broxson, E. H., Tucker, E. L., Giroux, D. S., Shults, J. Abshire, T. C. 1993. Prevalence of von Willebrand disease in children: A multiethnic study. *The Journal of pediatrics*, 123, 893-898.
- Westmoreland, D., Shaw, M., Grimes, W., et al. 2016. Super-resolution microscopy as a potential approach to diagnosis of platelet granule disorders. *Journal of Thrombosis and Haemostasis*, 14, 839-849.
- 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. *Journal of Thrombosis and Haemostasis*, 13, 815-820.

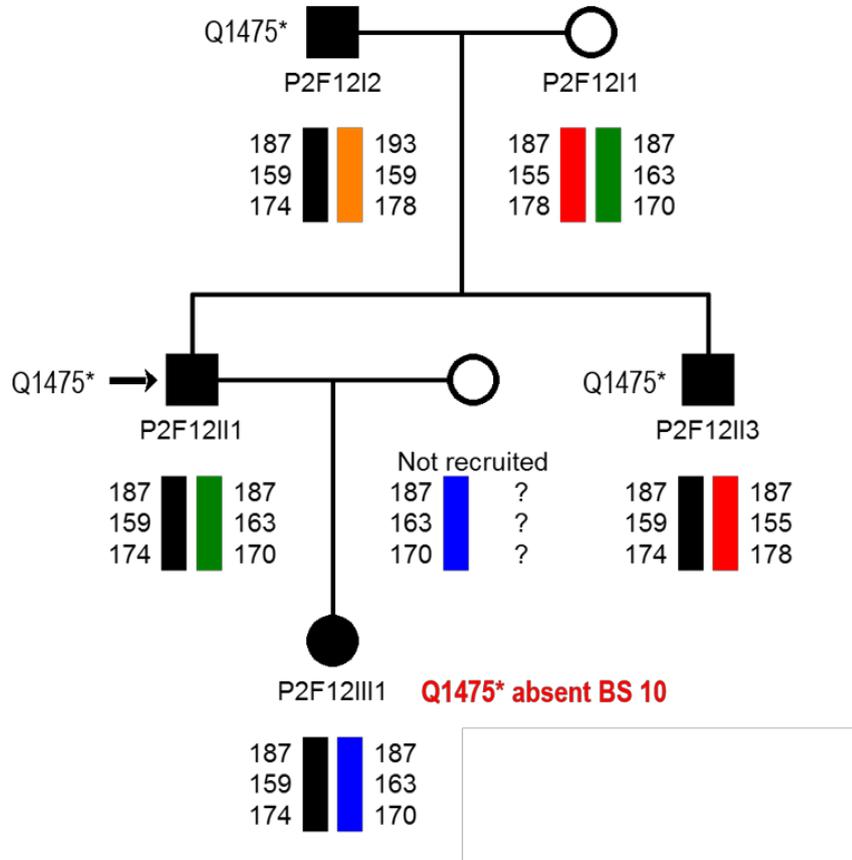
- Wyszynski, A., Hong, C.-C., Lam, K., et al. 2016. An intergenic risk locus containing an enhancer deletion in 2q35 modulates breast cancer risk by deregulating IGFBP5 expression. *Human Molecular Genetics*.
- Xie, F., Wang, X., Cooper, D. N., et al. 2006. A novel Alu-mediated 61-kb deletion of the von Willebrand factor (VWF) gene whose breakpoints co-locate with putative matrix attachment regions. *Blood Cells, Molecules, and Diseases*, 36, 385-391.
- Yadegari, H., Driesen, J., Hass, M., Budde, U., Pavlova, A. Oldenburg, J. 2011. Large deletions identified in patients with von Willebrand disease using multiple ligation-dependent probe amplification. *Journal of Thrombosis and Haemostasis*, 9, 1083-1086.
- Yang, L., Li, P., Fu, S., Calay, E. S. Hotamisligil, G. S. 2011. Defective Hepatic Autophagy in Obesity Promotes ER Stress and Causes Insulin Resistance. *Cell Metabolism*, 11, 467-478.
- Ye, S., Huang, Y., Joshi, S., et al. 2014. Platelet secretion and hemostasis require syntaxin-binding protein STXBP5. *The Journal of Clinical Investigation*, 124, 4517-4528.
- Yorimitsu, T., Nair, U., Yang, Z. Klionsky, D. J. 2006. Endoplasmic Reticulum Stress Triggers Autophagy. *Journal of Biological Chemistry*, 281, 30299-30304.
- Zarrei, M., Macdonald, J. R., Merico, D. Scherer, S. W. 2015. A copy number variation map of the human genome. *Nat Rev Genet*, 16, 172-183.
- Zhang, F., Gu, W., Hurles, M. E. Lupski, J. R. 2009. Copy Number Variation in Human Health, Disease, and Evolution. *Annual Review of Genomics and Human Genetics*, 10, 451-481.
- Zhang, Y., Zhang, M., Wu, J., Lei, G. Li, H. 2012. Transcriptional Regulation of the Ufm1 Conjugation System in Response to Disturbance of the Endoplasmic Reticulum Homeostasis and Inhibition of Vesicle Trafficking. *PLoS ONE*, 7, e48587.

- Zheng, X., Chung, D., Takayama, T. K., Majerus, E. M., Sadler, J. E. Fujikawa, K. 2001. Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura. *Journal of Biological Chemistry*, 276, 41059-41063.
- 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.
- Zhu, Q., Yamakuchi, M., Ture, S., et al. 2014. Syntaxin-binding protein STXBP5 inhibits endothelial exocytosis and promotes platelet secretion. *The Journal of Clinical Investigation*, 124, 4503-4516.
- Zimmerman, T. S., Ratnoff, O. D. Powell, A. E. 1971. Immunologic differentiation of classic hemophilia (factor VIII deficiency) and von Willebrand's disease: With observations on combined deficiencies of antihemophilic factor and proaccelerin (factor V) and on an acquired circulating anticoagulant against antihemophilic factor. *The Journal of Clinical Investigation*, 50, 244-254.

8 Appendices

8.1 Appendix 1

P2F12 family pedigree, laboratory phenotype and multimer profiles.



Center Code	Status	Bleeding Score	FVIII:C (IU/dl)	VWF:RCo (IU/dl)	VWF:Ag (IU/dl)	VWF:CB (IU/dl)	RCo/Ag	CB/Ag	RCo/CB	FVIII:C/VWF:Ag	Multimer Structure	VWF propeptide	VWFpp/VWF:Ag	Blood Group
P2F12 I:1	UFM	3	144	121	148	-	0.82	-	-	0.97	normal	134.9	0.9	O/O
P2F12 II:1	IC	3	36	38	27	48	1.41	1.78	0.79	1.33	normal	62.8	2.3	O/O
P2F12 I:2	AFM	-1	90	53	53	-	1	-	-	1.7	normal	92.1	1.7	O/O
P2F12 III:1	AFM	10	48	46	48	-	0.96	-	-	1	normal	86.3	1.8	O/O
P2F12 II:3	AFM	3	94	35	33	-	1.06	-	-	2.85	normal	77.2	2.3	O/O

8.2 Appendix 2

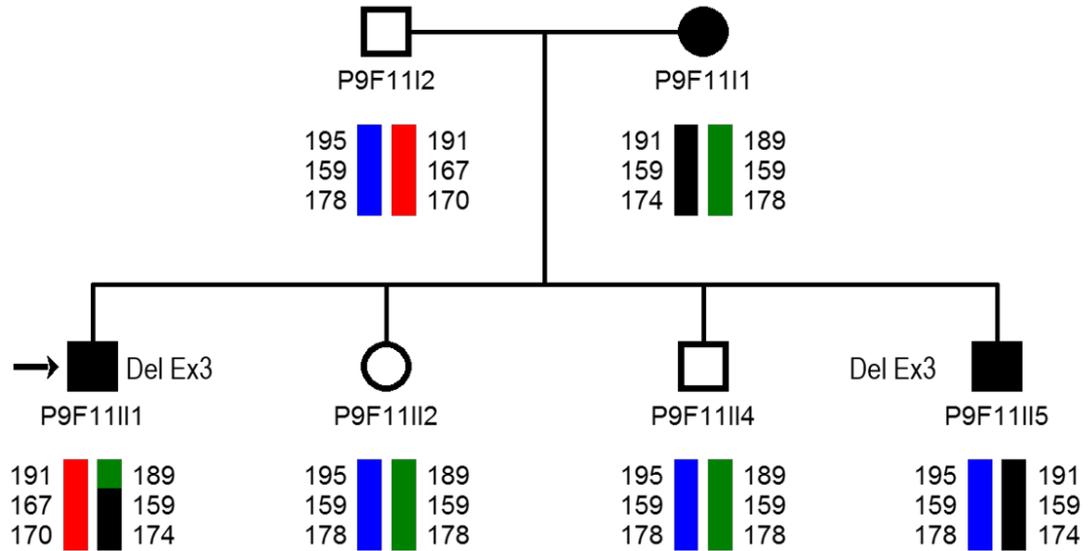
Breakdown of the 2504 samples from the 1000 genomes project into major population groups and their associated *VWF* 5' del frequencies.

Genotype	EUR (%) (n = 503)	AMR (%) (n = 347)	AFR (%) (n = 661)	EAS (%) (n = 504)	SAS (%) (N = 489)
5'del-I	11	1.4	ND	ND	4.7
5'del-II	11	4.6	1.2	0.2	1.8

EUR = European, AMR = Ad Mixed American, AFR = African, EAS = East Asian, SAS = South Asian
ND = Not Detected

8.3 Appendix 3

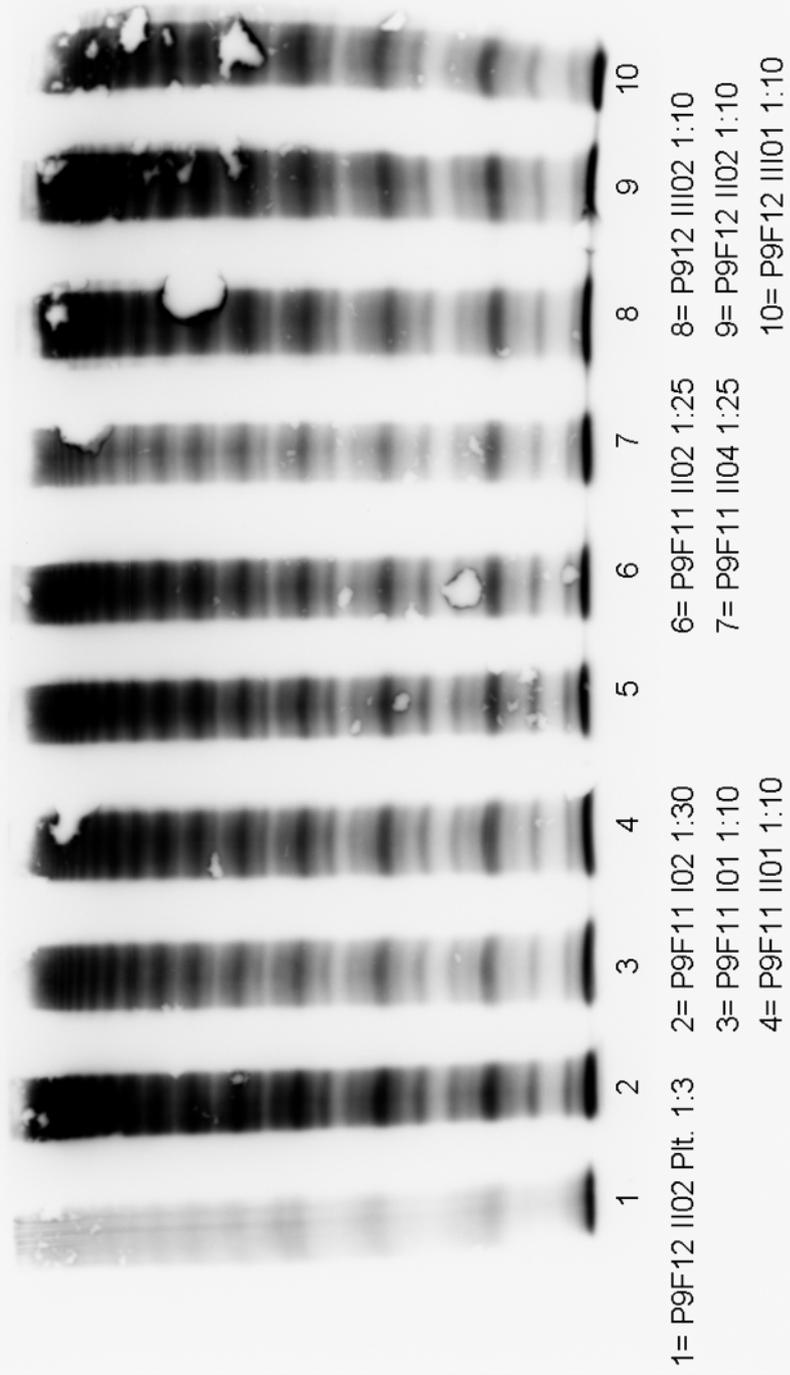
P9F11 family pedigree, laboratory phenotype and multimer profiles.



Family member	Status	Bleeding Score	FVIII:C (IU/dl)	VWF:RCo (IU/dl)	VWF:Ag (IU/dl)	VWF:CB (IU/dl)	RCo/Ag	CB/Ag	RCo/CB	FVIII:C/VWF:Ag	Multimer Structure	VWF propeptide	VWFpp/VWF:Ag	Blood Group
P9F11 I:2	UFM	-2	68	114	143	84	0.8	0.59	1.36	0.48	normal	139.8	1.0	O/A
P9F11 II:4	UFM	0	68	97	112	53	0.87	0.47	1.83	0.61	normal	124.2	1.1	A/A
P9F11 II:2	UFM	0	37	87	119	61	0.73	0.51	1.43	0.31	normal	103.3	0.9	A/A
P9F11 II:1	IC	10	5	20	31	31	0.65	1	0.65	0.16	normal	47.4	1.5	O/O
P9F11 I:1	AFM	10	21	28	35	25	0.8	0.71	1.12	0.6	normal	55.0	1.6	O/A
P9F11 II:5	AFM	6	16								NT		ND	A/A

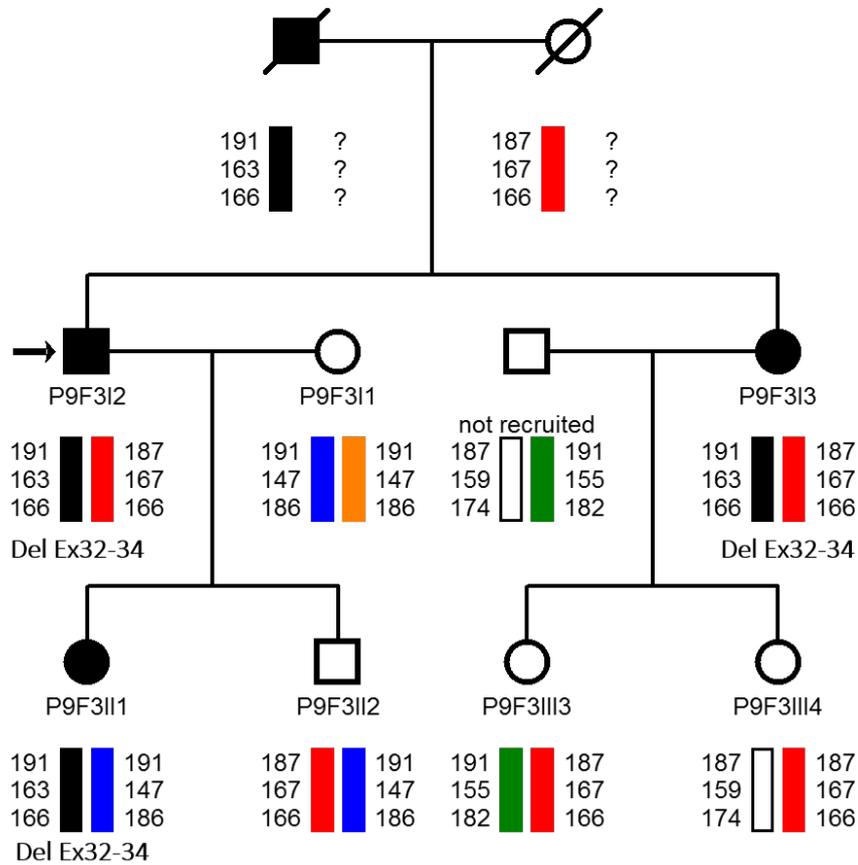
24.09.03 gel 1, medium resolution

5= NP 1:20



8.4 Appendix 4

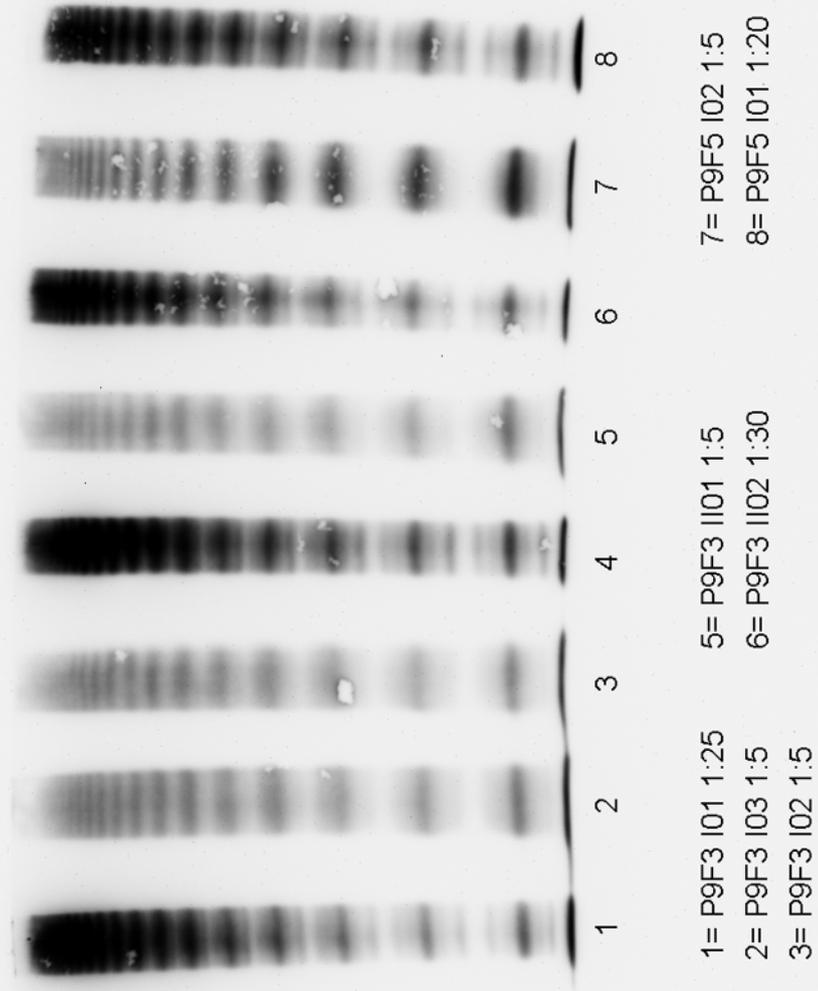
P9F3 family pedigree, laboratory phenotype and multimer profiles.



Family member	Status	Bleeding Score	FVIII:C (IU/dl)	VWF:RCo (IU/dl)	VWF:Ag (IU/dl)	VWF:CB (IU/dl)	RCo/Ag	CB/Ag	RCo/CB	FVIII:C/VWF:Ag	Multimer Structure	VWF propeptide	VWFpp/VWF:Ag	Blood Group
P9F3 II:2	UFM	0	65	99	122	-	0.81	-	-	0.53	normal	115.1	0.9	O/A
P9F3 I:1	UFM	0	72	89	105	-	0.85	-	-	0.69	normal	103.8	1.0	A/A
P9F3 I:2	IC	6	17	11	12	12	0.92	1	0.92	1.42	abnormal	57.1	4.8	O/A
P9F3 I:3	AFM	7	34	11	19	-	0.58	-	-	1.79	abnormal	64.6	3.4	O/A
P9F3 II:1	AFM	5	22	8	12	-	0.67	-	-	1.83	abnormal	48.8	4.1	A/A
P9F3 III:3	UFM	0	114	-	-	-	-	-	-	-	NT	-	-	ND A/A
P9F3 III:4	UFM	0	99	-	-	-	-	-	-	-	NT	-	-	ND O/O

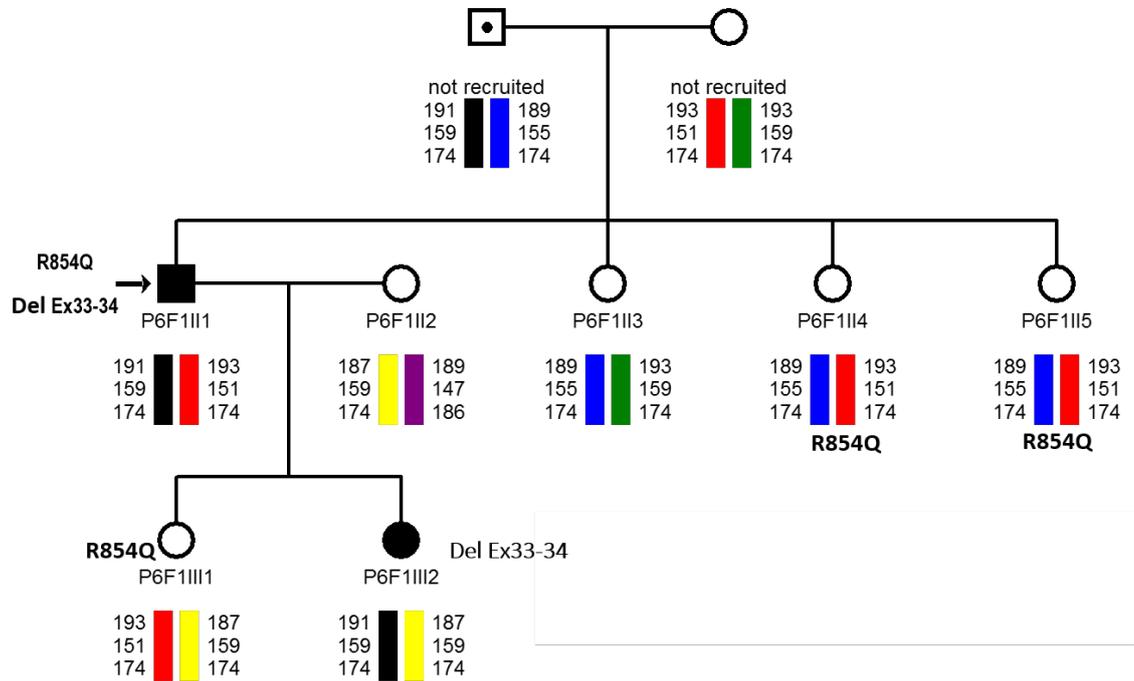
15.10.02 gel 3, medium resolution

4 = NP 1:20



8.5 Appendix 5

P6F1 family pedigree, laboratory phenotype and multimer profiles



Family member	Status	Bleeding Score	FVIII:C (IU/dl)	VWF:RCo (IU/dl)	VWF:Ag (IU/dl)	VWF:CB (IU/dl)	RCo/Ag	CB/Ag	RCo/CB	FVIII:C/VWF:Ag	Multimer Structure	VWF propeptide	VWFpp/VWF:Ag	Blood Group
P6F1 II:5	UFM	3	122	242	176	-	1.38	-	-	0.69	normal	166.9	0.9	O/A2
P6F1 II:4	UFM	4	89	224	145	-	1.54	-	-	0.61	normal	155.8	1.1	A/A2
P6F1 II:2	UFM	0	110	178	145	-	1.23	-	-	0.76	normal	132.8	0.9	O/O
P6F1 II:3	UFM	0	99	180	109	-	1.65	-	-	0.91	normal	126.1	1.2	A/A2
P6F1 III:2	AFM	8	29	14	21	-	0.67	-	-	1.38	abnormal	114.0	5.4	O/O
P6F1 III:1	UFM	5	59	114	73	-	1.56	-	-	0.81	normal	106.1	1.5	O/O
P6F1 II:1	IC	4	25	23	32	15	0.72	0.47	1.53	0.78	abnormal	130.8	4.1	O/O

11.9.02 gel 3, medium resolution

4 = NP 1:20

