Genetics of Atypical Haemolytic Uraemic Syndrome

By Rachel Challis

Thesis submitted in partial fulfilment of the requirements of the regulations for the degree of Doctor of Philosophy.

Institute of Genetic Medicine,

Newcastle University

October 2015

Abstract

Atypical haemolytic uraemic syndrome (aHUS) is a life threatening renal disease, caused by deregulation of the alternative complement pathway. Several genes within this pathway are associated with aHUS. At the outset of this project, the genetic cause had been identified in 45% of familial cases in the Newcastle aHUS cohort. The aim of this project was to identify the genetic cause of disease in the remaining 55%.

Complement Factor H (*CFH*) and Complement Factor H-related (*CFHRs*) are found on chromosome 1. This area contains several low copy repeats, the result of genomic duplications that occurred early in evolution. This causes genomic instability, which can lead to gene conversions or rearrangements. Sanger sequencing will not always detect these abnormalities, therefore patients were also screened using multiplex ligation-dependent probe amplification and western blotting.

A novel hybrid *CFH/CFHR3* gene was described, which arose by microhomology-mediated end joining. Functional analysis demonstrated that it was defective at regulating complement at the cell surface, which was predicted to predispose this patient to disease. Review of all patients in the Newcastle aHUS cohort with *CFH* abnormalities, identified a third of patients had a genomic rearrangement between *CFH* and *CFHRs*. The relative frequency of genomic rearrangements emphasised the importance of undertaking copy number analysis in aHUS diagnostic testing, because often they are not detected by Sanger sequencing.

Whole exome sequencing was then undertaken in Newcastle familial cohort with an unknown genetic aetiology. Pathogenic sequence variants were identified in genes, known to be associated with thrombotic microangiopathies. Sequence variants that were predicted to be pathogenic, were found in three genes not previously associated with disease. Two of these genes were located outside of the complement system, indicating that complement-directed therapies may be contraindicated. In this project, a genetic cause of disease was found in 54% of familial cases tested.

Acknowledgements

I would like to take this opportunity to thank everyone who has helped me over the past three years and contributed to the work demonstrated here.

Firstly I wish to thank my supervisors David Kavanagh and Tim Goodship for giving me the opportunity to carry out this work and for their guidance and support throughout. I want to thank David especially for his mentorship, which has been beyond what any student could hope for.

I want to thank Holly Anderson and Edwin Wong for all their help and advice, it has been a pleasure to work with you both. I would like to thank Kevin Marchbank for all the advice he has given. I wish to thank my PhD progress panel members John Sayer and Neil Sheerin for their time and recommendations throughout the three years. I would also like to thank Lisa Turnbull and Valerie Wilson for all their help.

I would like to take the opportunity to acknowledge the people who have also contributed to the work that has been carried out in this project. Firstly to Achim Truemann for carrying out mass spectrometry for this project. Then to Claire Harris and Paul Morgan for kindly providing reagents. Yaobo Xu and Mauro Santibanez-Koref for their work on the WES bioinformatics, a critical component of this project. Finally to Geisilaine Soares dos Reis Araujo for her contribution to the patient screening carried out here.

I want to thank my friends in Newcastle who have made the last 3 years really wonderful, I will be sad to go. I also wish to thank my amazing friends Audrey, Frances, James and Emma for their support throughout the past three years. I would also like to thank Alberto Briones-León for his support and for providing me with a continuous supply of Nutella.

Finally I want to thank my parents Heather and Jonathan and my brother Richard. I wouldn’t have been able to do this without your endless support and encouragement. I love you and dedicate this to you.

Contents

List of Figures

List of Tables

List of Abbreviations

°C	Degrees Celsius
1000g	1000 Genomes Project
α'NT	N-terminus of cleaved α -chain
aCGH	Array comparative genomic hybridisation
ADP	Adenosine diphosphate
aHUS	Atypical Haemolytic Uraemic Syndrome
AMD	Age-related Macular degeneration
AP	Alternative Pathway
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
C3/4/5/6/7/8/9	Complement protein 3/4/5/6/7/8/9
C4BP	C4b binding protein
Ca^{2+}	Calcium ions
Cbl	Cobalamin
cDNA	Complementary DNA
CFD	Complement Fixation Diluent
CFB	Complement Factor B gene
CFH	Complement Factor H gene
CFHR	Factor H related gene
CFI	Complement Factor I gene
CMT	Charcot Marie Tooth disease
CNS	Central Nervous System
COX	Cyclooxygenase
CP	Classical Pathway
CCPs	Complement Control Proteins
CD46	Membrane Cofactor Protein
CMT	Charcot Marie Tooth
CNV	Copy number variation
$cPLA_2$	Cytosolic phospholipase A_2
CR	Complement Receptor
Cys1	Cysteine-rich domain
DAD	Diaphanous autoregulatory domain
DAF	Decay Accelerating Factor
DAG	Diacylglycerol
dbSNP	Single Nucleotide Polymorphism Database
DGKs	Diacylglycerol kinases
DGKε and *DGKE*	Diacylglycerol Kinase ε
DID	Diaphanous Inhibitory Domain
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DSB	Double strand break
EA	Sensitised sheep erythrocytes
ECM	Extracellular matrix
EGF	Epidermal growth factor
ER	Endoplasmic reticulum
ESP6500	NHLBI GO Exome Sequencing Project
ESRF	End Stage Renal Failure
F12	Coagulation factor XII
FATHMM	Functional Analysis through Hidden Markov Models
FB	Factor B protein
FD	Factor D protein
FDP	Fibrin Degradation Products

FFPE	Formalin fixed paraffin embedded
FH	Factor H protein
FHL	Factor H-like
FHR	Factor H related protein
FH1/2	Formin Homology 1/2
FHR	Complement Factor H Related
FI	Factor I protein
FKRP	Fukutin related protein
FSGS	Focal Segmental Glomerulosclerosis
GAGs	Glycolaminoglycans
GATK	Genome Analysis Toolkit
GBD	GTPase-binding domain
GBM	Glomerular basement membrane
GC	Guanine-Cytosine
GFR	Glomerular filtration rate
gDNA	Genomic DNA
GPI	Glycosylphosphatidylinositol
GWAS	Genome- Wide Association Study
H_2O	Water
Hb	Haemoglobin
HD	Hydrophobic domain
HE	Haematoxylin and eosin staining
HIV	Human immunodeficiency virus
HJ	Holliday Junctions
HMECs	Human dermal microvascular endothelial cells
HR	Homologous recombination
HUS	Haemolytic Uraemic Syndrome
HUVEC	Human umbilical vein endothelial cells
IGV	Integrative Genomics Viewer
INDEL	Insertion/deletion
INF2	Inverted Formin 2
IL	Interleukin
IVA	Ingenuity Variant Analysis
KAD	Kinase accessory domain
kb	Kilo base pairs
KCD	Kinase catalytic domain
kDa	Kilo Dalton
LCR	Low copy repeats
LDH	Lactate dehydrogenase
LDL	Low-density Lipoprotein
LP	Lectin Pathway
MAC	Membrane Attack Complex
MAF	Minor Allele Frequency
MAHA	Microangiopathic haemolytic anaemia
MAL	Myelin and lymphocyte protein
MAPK	p38 Mitogen-activated protein kinase
MASP	MBL- associated serine proteases
MBL	Mannose-Binding Lectin
MCP	Membrane cofactor protein
MEK	MAPK kinase
Mg^{2+}	Magnesium ion
MGN	Membranous glomerulonephritis
MHC	Major Histocompatibility Complex
MMACHC	Methylmalonic aciduria *cblC* type with homocystinuria

MMEJ	Microhomology- mediated End Joining
MPGN	Mebranoproliferative glomerulonephritis
mRNA	Messenger RNA
NAHR	Non-allelic homologous recombination
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next Generation Sequencing
NHEJ	Non-homologous end joining
NHS	N-hydroxysuccinimide
NHSΔBΔH	Normal human serum depleted of FB and FH
OmCI	*Ornithodoros moubata* complement inhibitor
PA	Phosphatidic acid
PAS	Periodic acid–Schiff staining
PAI-1	Plasminogen activator inhibitor-1
PBMCs	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PDGF	Platelet-derived growth factor
PI	Phosphatidylinositol
PIP_2	Phosphatidylinositol (4,5)-bisphosphate
PKC	Protein Kinase C
$PLA_2\gamma$	Phospholipase A_2 gamma
PLC	Phospholipase C
PLG	Plasminogen
PNH	Paroxysmal Nocturnal Haemoglobinuria
QC	Quality control
RCA	Regulators of Complement Activation
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute
RR	Genotype for homozygous reference allele
RV	Genotype for heterozygous variant allele
SBS	Sequencing by Synthesis
SD	Standard deviation
SDS PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SDSA	Synthesis-dependent strand-annealing
SLE	Systemic lupus erythematosus
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
ssDNA	Single-stranded DNA
STEC	Shiga toxin Escherichia coli
Stx	Shiga toxin
STPR	Serine-threonine-proline rich
TTP	Thrombotic thrombocytopenia purpura
THBD	Thrombomodulin
TMA	Thrombotic microangiopathy
TRAF6	TNF receptor-associated factor 6
TRPC6	Transient receptor potential cation channel, subfamily C, member 6
UTR	Untranslated region
VEGF	Vascular endothelial growth factor
VEGFR	VEGF receptor
v/v	Volume/volume
vWA	Von Willebrand A
vWF or *VWF*	von Willebrand factor

VV	Genotype for homozygous variant allele
w/v	Weight/volume
WES	Whole Exome Sequencing
WGA	Whole Genome Amplification
WGS	Whole Genome Sequencing
WH2	Wiskott-Aldrich syndrome protein-homology 2

Chapter 1: Introduction

1.1. Haemolytic uraemic syndrome

Haemolytic uraemic syndrome (HUS) is a thrombotic microangiopathy (TMA) that is characterised by the occurrence of microangiopathic haemolytic anaemia (MAHA), thrombocytopenia and acute renal failure in patients (Siegler and Oakes, 2005, Kavanagh and Goodship, 2010, Fremeaux-Bacchi *et al.*, 2008, Boyer and Niaudet, 2011). MAHA refers to the lysis of erythrocytes, which occurs as a result of increased shear stress in diseased blood vessels (Barbour *et al.*, 2012). Thrombocytopenia describes the decrease in the amount of platelets found in the blood due to their sequestration within blood clots. In this condition the kidney is unable to filter the blood efficiently, as a result of the thrombi in the microvasculature and consequently, there is ischaemia and reduced kidney function (Coppo and Veyradier, 2009). There are two subtypes of HUS, the typical and atypical form.

1.1.1. Typical HUS

Typical HUS or Shiga toxin *Escherichia coli* (STEC) HUS is caused by the presence of Shiga toxin (Stx)- producing bacteria (Kavanagh *et al.*, 2014). It is the most commonly occurring form of HUS, comprising of approximately 90% of cases (Boyer and Niaudet, 2011). It occurs at a frequency of approximately 2 to 3 per 100,000 in the population, most frequently in young children (Tarr *et al.*, 2005). One commonly HUS-associated serotype of STEC is the O157:H7 strain (Tarr *et al.*, 2005), although there are other Shiga toxin-producing bacteria that can cause HUS such as *Shigella dysenteriae* serotype 1 (Koster *et al.*, 1978). Patients with STEC HUS often present with bloody diarrhoea (Barbour *et al.*, 2012).

1.1.2. Atypical HUS

Atypical or non-diarrhoeal HUS (aHUS) is a rare form of HUS, found in the remaining 10% of cases. aHUS occurs at a frequency of approximately 7 per million in children and 2 per million in adults (Taylor *et al.*, 2010). This subtype of HUS is defined by the absence of STEC in patient stool samples (Scully and Goodship, 2014). It is a severe disease with 70% of patients developing End Stage Renal Failure (ESRF) and a 10-15% mortality rate (NICE, 2015).

A patient with aHUS has a reduced ability to filter waste from the blood due to pathology in glomerulus (Kavanagh *et al.*, 2013). The glomerulus is a dense network of capillaries that form part of the nephron, the filtering unit of the kidney (Pollak *et al.*, 2014). There are approximately 1 million nephrons per kidney (Bertram *et al.*, 2011). In aHUS the glomerular capillaries become occluded causing a loss of kidney function (Pollak *et al.*, 2014). This is the result of narrowing of the glomerular capillaries, as a result of endotheliosis (endothelial cell swelling), deposition of material between the endothelium and glomerular basement membrane (GBM) and the formation of fibrin and platelet-rich thrombi (Kavanagh and Goodship, 2010). Figure 1-1 demonstrates the occlusion of the glomerular capillaries in a histological section of a glomerulus from a patient with aHUS.

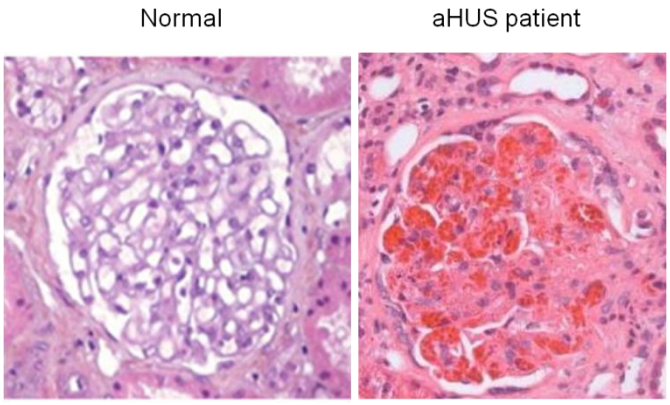

Figure 1-1 Histological sections of glomeruli.

Showing normal histology (Kamar et al., 2008) (left) and aHUS histology (McMahon, 2011) (right).

It has now been well documented that loss of complement regulation and thus uncontrolled alternative pathway (AP) activation is the initiating event that leads to aHUS (Meri, 2013). This includes the occurrence of activating and inactivating mutations in genes that encode proteins found within the complement pathway (Kavanagh *et al.*, 2013, Fremeaux-Bacchi *et al.*, 2013, Maga *et al.*, 2010, Noris *et al.*, 2010). The inheritance pattern of aHUS has been previously reported to be both recessive (Thompson and Winterborn, 1981) and dominant (Roodhooft *et al.*, 1990).

1.1.3. Diagnosis

HUS is diagnosed by the findings of MAHA, thrombocytopenia and acute renal failure, shown in Table 1.

Test	HUS
Haemoglobin (Hb)	Reduced
Platelet count	Reduced
Haptoglobin	Reduced
Blood smear	Schistocytes
Coombs test	Negative*
Serum creatinine	Elevated
Lactate dehydrogenase (LDH)	Elevated

Table 1 Diagnostic tests for HUS.

'' Pneumococcal HUS positive.*

Originally aHUS was used to classify any HUS in the absence of Stx-producing bacteria (Kavanagh *et al.*, 2013). Primary aHUS had been used to refer to cases with complement deregulation caused by complement gene mutations and secondary aHUS was used to describe the development of aHUS after a trigger event (Besbas *et al.*, 2006), shown in Table 2. However this failed to account for the fact that often, patients with an underlying complement mutation will still often require a secondary trigger for the disease to manifest. Classifications that consider the genetic and environmental causes of disease are now being introduced (Besbas *et al.*, 2006).

Event	Reference
Infections	Cochran *et al.* (2004), Constantinescu *et al.* (2004)
Drugs	Al-Nouri *et al.* (2015), Eremina *et al.* (2008), Zarifian *et al.* (1999)
Autoimmune conditions	El-Husseini *et al.* (2015), Aguiar and Erkan (2013)
Transplants	Noris and Remuzzi (2010)
Pregnancy	Fakhouri *et al.* (2010)
Metabolic conditions	Sharma *et al.* (2007)

Table 2 Events that can lead to aHUS.

Reviewed by Kavanagh et al. (2013).

1.2. Complement System

The complement system is part of the innate immune system, which is involved in producing a large, non-specific immune reaction in response to foreign microorganisms (Trouw and Daha, 2011). It is composed of three pathways: the classical (CP), lectin (LP) and AP, which are initiated by different factors (Campbell *et al.*, 1988). Ultimately the pathways combine to produce a common immunological response (Walport, 2001a, Walport, 2001b). This includes the release of anaphylatoxins which recruit other

immunological cells, leading to amplification of the immune response and the formation of a Membrane Attack Complex (MAC) that can destroy targeted cells (Sarma and Ward, 2011).

The AP is unlike the CP and LP as it does not require an initiating step. Instead it depends on the natural hydrolysis of a C3 thioester, producing C3(H_2O) (Thomas *et al.*, 1982). C3(H_2O) is structurally similar to C3b thus it is able to bind with complement factor B (FB) (Isaac and Isenman, 1992). However the CP and LP can also recruit the AP, leading to pathway amplification. The AP C3 convertase, a serine protease, is formed by a C3b subunit binding to FB to form a pro-convertase enzyme (Zipfel and Skerka, 2001). FB contains the serine protease domain of the C3 convertase, which is not activated until complement factor D (FD), another serine protease, cleaves FB into a Ba and Bb subunits (Sarma and Ward, 2011). The Bb subunit containing the active region remains bound to the C3b, producing the active AP C3 convertase (C3bBb complex), whilst the anaphylatoxin, Ba subunit, is released (Ricklin *et al.*, 2010). With the additional stabilisation effect of Properdin, the C3 convertase is then able to cleave more C3 into C3a and C3b. This leads to generation of even more C3 convertase, producing a positive feedback loop (Holers, 2014), shown in Figure 1-2. Some C3b subunits can complex with the C3 convertase to produce a C5 convertase, which can cleave C5 into C5a and C5b (Campbell *et al.*, 1988). C5a is also an anaphylatoxin that is released into the circulation whilst C5b forms a complex with complement components 6 to 9 to produce a MAC (Hadders *et al.*, 2012). This is able to insert into the plasma membrane of pathogens or self cells, leading to lysis of targeted cells (Ricklin *et al.*, 2010).

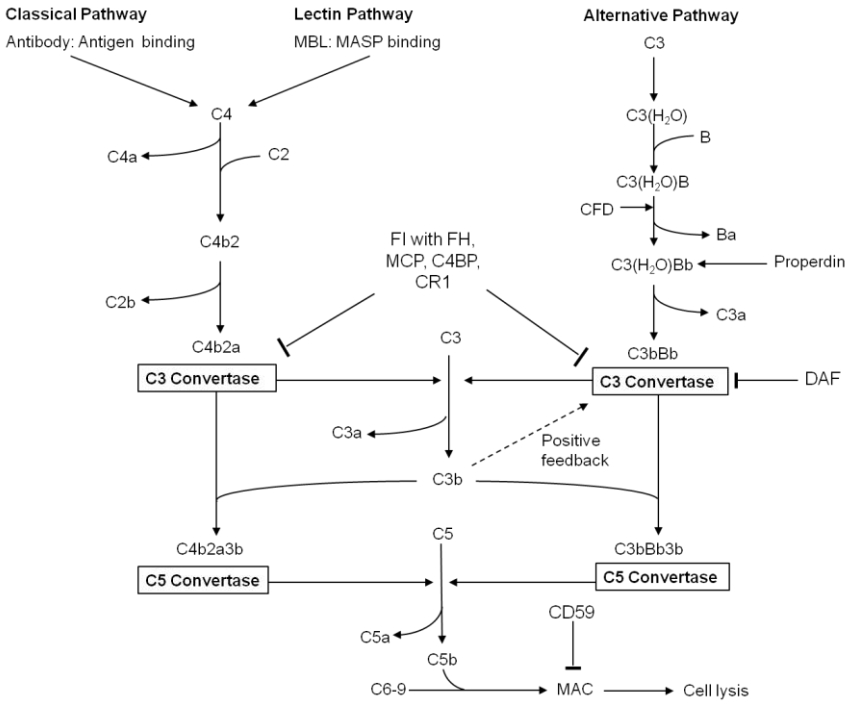

Figure 1-2 Schematic diagram of the Complement system.

Adapted from Campbell et al. (1988).

1.2.1. Complement Regulators

Regulators of complement activation are molecules that control the complement system when it is not required. They include Factor H (FH), Factor I (FI), Complement Receptor 1 (CR1), C4b binding protein (C4BP), Membrane Cofactor Protein (MCP), CD59 and Decay Accelerating Factor (DAF) (Holers, 2014). There are three methods for complement regulation including protein mediated decay acceleration, irreversible protein inactivation and reversible competitive inhibition (Hourcade, 1989).

Firstly protein-mediated decay acceleration occurs when the C3 or C5 convertase enzymes dissociate more readily, due to the presence of regulators such as FH (Schmidt *et al.*, 2008a) and DAF (Sun *et al.*, 1999). The second mechanism is irreversible inactivation, which occurs when a protein is cleaved into smaller non-functioning by-products (Hourcade, 1989). For example C3b is first cleaved by FI, with fluid phase cofactors FH or cell surface cofactors MCP and CR1, to produce a surface bound iC3b and a soluble C3f subunit (Davis *et al.*, 1984). C4BP predominately regulates the CP, however it has also been shown to act as a cofactor to FI in the inactivation of C3b (Blom *et al.*, 2003). The second proteolytic event uses CR1 as a cofactor to cleave iC3b into a membrane-bound C3dg and a soluble C3c subunit (Davis *et al.*, 1984, Hourcade,

1989). Finally competitive inhibition is the reversible binding of regulators to target molecules, so that they are unable to bind to other molecules and carry out their function. For example FH can competitively bind with C3b, reducing the amount available for FB, for C3 convertase formation (Schmidt *et al.*, 2008a).

1.3. Known genetic causes of aHUS

1.3.1. Complement Factor H (CFH)

The Complement Factor H (*CFH)* is a gene containing 23 exons, located in the Regulators of Complement Activation (RCA) cluster on Chromosome 1q32. It encodes FH, a 155kD protein (Morgan *et al.*, 2011) that contains 20 Complement Control Protein domains (CCPs) in a folded conformation (Morgan *et al.*, 2011), shown in Figure 1-4. FH is a serum glycoprotein produced mainly by the liver, which regulates the alternative pathway by several mechanisms. Firstly it competes for the binding site on C3b with FB and C5, thus restricting the formation of C3 convertases (C3bBb complex) and C5 convertases (C3BbC5a) (Warwicker *et al.*, 1998). The second method is to act as a cofactor to FI in order to mediate the proteolysis and thus inactivation of C3b (DiScipio, 1992). The final function of FH is to enhance the dissociation of the AP C3 convertase by releasing the C3b and Bb subunits (Hourcade *et al.*, 2002).

Functional analysis of FH has demonstrated that there are 3 regions involved in C3b binding. These areas are CCPs 1-4, CCPs 6-8 and CCPs 19-20 (Sharma and Pangburn, 1996, DiScipio, 1992, Warwicker *et al.*, 1998), although CCPs 1-4 and 19-20 are predominantly involved (Schmidt *et al.*, 2008b). FH also has the ability to bind to polyanions, such as glycolaminoglycans (GAGs) found on cell membranes, via CCPs 6-8, 19 and 20 (Ferreira *et al.*, 2009, Blaum *et al.*, 2015, Blackmore *et al.*, 1996). One study demonstrated that the removal of GAGs from sheep erythrocytes led to impaired complement regulation due to reduced FH-activity (Fearon, 1978). CCPs 1-4 of FH have also been demonstrated to mediate cofactor (Gordon *et al.*, 1995, Kuhn *et al.*, 1995, Sharma and Pangburn, 1996) and decay acceleration activity of FH (Kuhn and Zipfel, 1996). Alternative splicing of *CFH* produces a truncated protein termed FH-like (FHL), which contains CCPs 1-7 (Kuhn *et al.*, 1995). The role of FHL is not currently known, although the presence of CCPs1-4, suggests that it also binds to C3b and mediates

cofactor activity with FI (Zipfel and Skerka, 1999). These binding sites are shown on the FH protein in Figure 1-3.

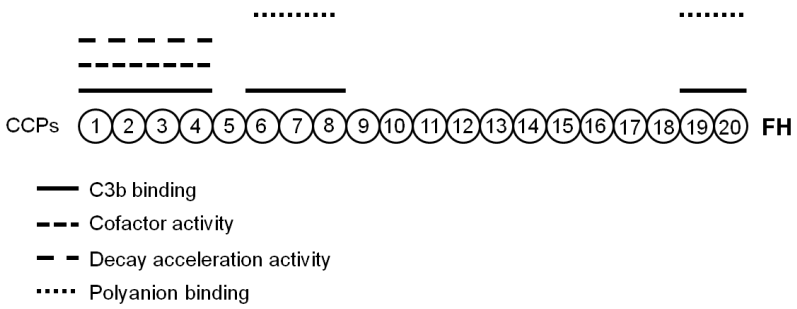

Figure 1-3 Map of FH binding sites

Adapted from Schmidt et al. (2008b).

CFH was the first gene to be linked to aHUS (Warwicker *et al.*, 1998), mainly inherited in an autosomal dominant mode of inheritance. *CFH* mutations currently are found in approximately 30% of aHUS cases (Caprioli *et al.*, 2006, Loirat and Fremeaux-Bacchi, 2011, Maga *et al.*, 2010). As a result of these inactivating mutations, there is reduced AP control and thus onset of disease. Patients with *CFH* mutations may have lower C3 levels due to uncontrolled AP pathway and thus increased C3 consumption (Kavanagh *et al.*, 2013). Patients with mutations in *CFH* have a poor prognosis, with approximately 60% of cases resulting in death or ESRF (Loirat *et al.*, 2008, Sellier-Leclerc *et al.*, 2007). In addition patients with *CFH* mutations are more likely to have recurrence of disease after transplantation, with approximately 80% of patients losing the graft within 2 years (Bresin *et al.*, 2006, Richards *et al.*, 2001). This is due to the presence of FH in the circulation, therefore a renal allograft will not correct the underlying genetic defect.

It has been observed that many disease-associated mutations are located in the C-terminal portion of the protein, particularly within CCPs domains 19 and 20 (Noris and Remuzzi, 2009, Caprioli *et al.*, 2001, Perez-Caballero *et al.*, 2001, Kavanagh *et al.*, 2013), shown in Figure 1-4. This area is highly important in cell surface binding, therefore mutations occurring here can lead to reduced FH-mediated complement regulation on the cell surface (Ferreira *et al.*, 2006, Ferreira *et al.*, 2009). Mouse models have shown that mice lacking CCPs 16-20 (FHΔ16-20) developed aHUS (Pickering *et al.*, 2007). Goicoechea de Jorge *et al.* (2011) crossed FHΔ16-20 mice with C5 knockout mice and found that these mice failed to develop aHUS spontaneously. This indicated the importance of the terminal pathway in disease pathogenesis.

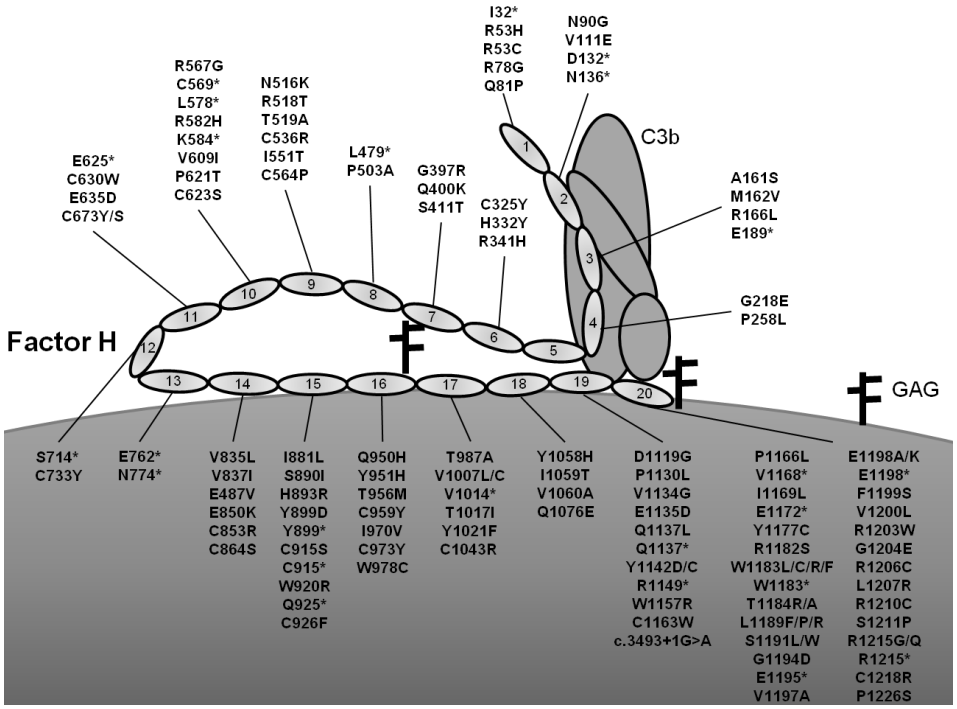

Figure 1-4 FH protein and mutations

Schematic diagram showing FH on the cell surface, annotated with point mutations seen in aHUS. Adapted from Kavanagh et al. (2013), variants obtained from the literature and from FH HUS database (Rodriguez et al., 2014).

1.3.1.1. CFH risk haplotype

Additional sequence variants have been found in *CFH* that do not directly cause aHUS, but are thought to give carriers a higher risk of developing aHUS (Pickering *et al.*, 2007). This haplotype is known as *CFH*-H3, shown in Table 3.

SNP	Amino acid change	dbSNP ID	MAF (%)
-331C>T	-	rs3753394	26.2
c.184G>A	V62I	rs800292	46.8
c.1204C>T	Y402H	rs1061170	26.7
c.2016A>G	Q672=	rs3753396	20.3
c.2237-543G>A	-	rs1410996	49.5
c.2808G>T	E936D	rs1065489	20.3

Table 3 CFH - H3 aHUS risk SNPs.

Minor allele frequency (MAF) is given as a percentage, based on the 1000g database.

Caprioli *et al.* (2003) demonstrated that -331C>T, c.2016A>G and c.2808G>T were significantly (c.1204C>T mildly) associated with aHUS in patients, with and without additional *CFH* mutations. However when investigated in the Newcastle cohort, this association was only observed in aHUS patients without additional *CFH*, *MCP* or *CFI*

sequence variants. In the French cohort an association was found in patients with known mutations (in *CFH*, *MCP* or *CFI*) (Fremeaux-Bacchi *et al.*, 2005). These differences may be the result of screening small populations. Heurich *et al.* (2011) and Hocking *et al.* (2008) demonstrated that the polymorphism V62 decreased FH's affinity to C3b, which would cause less C3b inactivation.

1.3.2. CFH related (CFHRs) genes

Within the RCA cluster there are several genes upstream of *CFH*, called Complement Factor H related proteins (*CFHRs*). The *CFHRs* are thought to have arisen during evolution in a series of sequence duplication events (Jozsi and Zipfel, 2008). As a result there are areas of very high sequence similarity, called low copy repeats (LCRs) between *CFH* and *CFHRs* (Bailey *et al.*, 2002), shown in Figure 1-5.

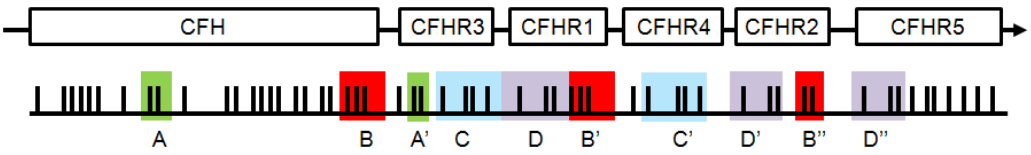

Figure 1-5 Position of CFH and CFHRs on Chromosome 1.

A diagram showing region of Chromosome 1 that encodes CFH and CFHRs. LCRs are colour coded and labelled A-D and exons are shown as vertical lines. Adapted from Heinen et al. (2006).

Currently there have been 5 FHRs described, labelled 1-5. They share similar features to FH in that they are mainly secreted by the liver and they are composed of CCPs domains. Figure 1-6 shows the percentage amino acid homology of the FHRs compared to FH. Overall, the CCPs of all 5 FHRs range from approximately 30-100% sequence identity to FH. FHR1 has the highest sequence identity to FH, particularly CCPs 3-5, which are between 95-100% similar to CCPs 18-20 of FH. Due to the lack of protein domains that correspond to CCPs 1-4 of FH, it is expected that the FHRs do not display cofactor or decay acceleration activity (Jozsi and Zipfel, 2008). This is true for all FHRs except for FHR3, 4 and 5 which display some cofactor activity (Hellwage *et al.*, 1999, McRae *et al.*, 2005). The high sequence identity of the FHRs 1,2 and 5 N-terminal CCPs allowed them to dimerise and compete with FH, leading to reduced complement regulation at the cell surface (Goicoechea de Jorge *et al.*, 2013). Although alternative hypotheses have suggested that they may have a protective effect by regulating complement in addition to FH (Jozsi and Zipfel, 2008).

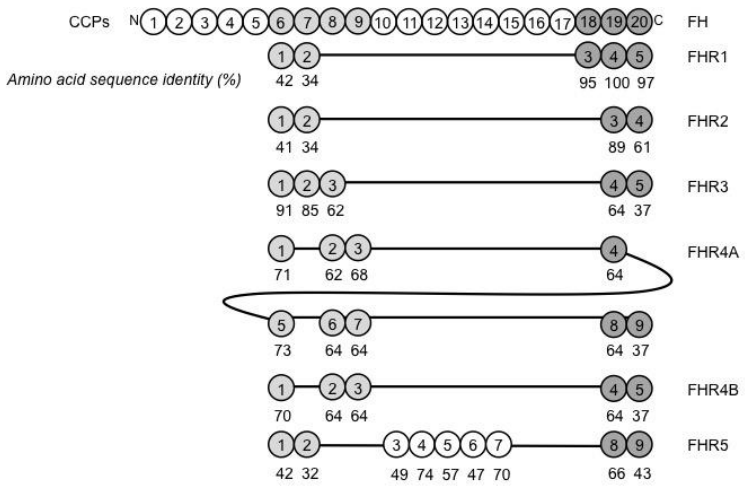

Figure 1-6 FH and FHRs amino acid sequence identity.

Light grey indicates the CCPs that are homologous to CCPs 6-9 of FH and dark grey indicates CCPs that are homologous to CCPs 18-20 of FH. Amino acid identity to FH is given as a percentage. Adapted from Jozsi and Zipfel (2008).

Studies have shown that copy number variation (CNV) of *CFHRs*, which is the loss or gain of approximately ≥1kb genetic material (Chen *et al.*, 2010), is associated with the presence of anti-FH autoantibodies that bind to and inhibit the C-terminus of FH (Dragon-Durey *et al.*, 2009, Zipfel *et al.*, 2007, Moore *et al.*, 2010, Jozsi *et al.*, 2008, Abarrategui-Garrido *et al.*, 2009). These studies identified that homozygous *CFHR1/3* or *CFHR1/4* deficiency was higher in aHUS patients than in control populations, especially in those patients with anti-FH autoantibodies (Zipfel *et al.*, 2007, Jozsi *et al.*, 2008, Dragon-Durey *et al.*, 2009, Abarrategui-Garrido *et al.*, 2009, Moore *et al.*, 2010). It is unclear whether it is the deletion and the loss of FHR protein that alters complement regulation and that the autoantibody is an epiphenomenon or that it is the deletion that allows the development of autoantibodies and these cause disease. CNVs can occur as a result of structural rearrangements in the DNA, discussed below.

1.3.2.1. DNA repair mechanisms

1.3.2.1.1. Homologous Recombination (HR)

When a double strand break (DSB) has occurred during meiosis, there are several mechanisms used to repair the DNA. Homologous recombination (HR) is one that utilizes homologous sequences to act as a template, allowing polymerases to restore the missing sequences(McVey and Lee, 2008). Here the 5' end of the strands are shortened to form single-stranded DNA (Szostak *et al.*, 1983). These hybridise with a nearby

sequence displaying a high degree of homology, creating a displacement (D)-loop where they are then extended by DNA synthesis. The other 3' single-stranded DNA (ssDNA) undergoes DNA synthesis, using the D-loop as a template sequence (Chen *et al.,* 2010). This can lead to non-allelic homologous recombination (NAHR) or gene conversions, which can cause CNVs. It can lead to the formation of hybrid genes, several of which have been reported in association with aHUS (Heinen *et al.,* 2006, Venables *et al.,* 2006, Eyler *et al.,* 2013, Valoti *et al.,* 2015, Francis *et al.,* 2012). These hybrid genes will be discussed in Chapter 3.

Once the D-loop has formed there are two possible pathways that lead to either NAHR or gene conversion. The first uses the Holliday Junctions (HJ) model, where DNA synthesis of the 5' strand occurs and double HJs are formed due to ligation of the strand break (Holliday, 2007). HJs are intermediate structures occurring between all 4 strands, which are then cleaved by endonucleases (Chen *et al.,* 2010). This can occur in two possible orientations, reviewed by Liu and West (2004). It can occur on the strand that is complementary to the strand break, indicated by the solid grey arrows in Figure 1-7, which leads to crossing over and thus NAHR. The alternative mechanism is that the other set of strands are cleaved, shown by the hashed grey arrows, where there is no crossing over and leads to a gene conversion.

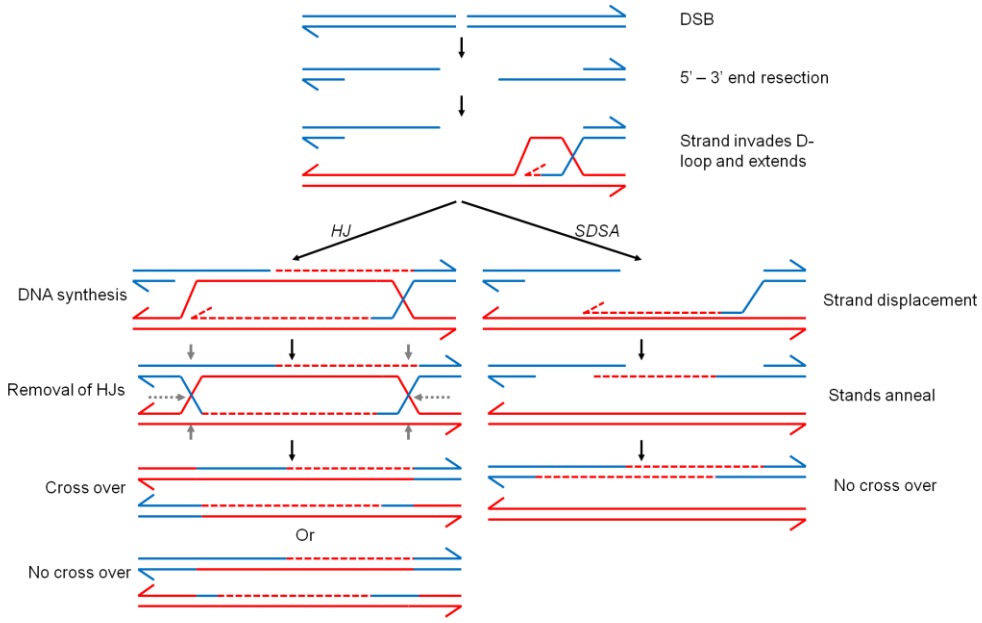

Figure 1-7 Schematic diagram showing DSB repair by HR.

DSB repair starts with a common mechanism, before using either HJ or SDSA repair model. HJ leads to both NAHR and gene conversions, whereas SDSA leads only to gene conversions. Adapted from Szostak et al.(1983) and Chen et al. (2007).

Gene conversions can also occur via another mechanism, termed the synthesis-dependent strand-annealing (SDSA) pathway, where after D-loop formation and 5' extension, the strand becomes displaced (Chen *et al.*, 2007). It then anneals to the 3'strand, where DNA synthesis continues, shown in Figure 1-7.

1.3.2.1.2. Microhomology-mediated End Joining (MMEJ)

Microhomology-mediated end joining (MMEJ) is a distinct subtype of non-homologous end joining (NHEJ). Unlike HR, NHEJ mechanisms repair DSBs without the use of homologous sequences, with the exception of MMEJ (McVey and Lee, 2008). MMEJ take places less frequently, occurring when microhomologies of 5-25bp are found next to DSBs (Sharma *et al.*, 2015). Similar to HR, after the occurrence of a DSB the 5' ends are resected to reveal microhomologous sequences that then align and anneal. DNA synthesis and ligation then restores the gaps in the sequences.

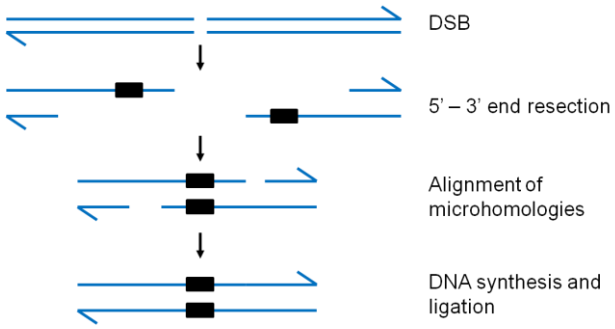

Figure 1-8 Schematic diagram showing DSB repair by MMEJ

Blue lines represent DNA strands and black squares demonstrate regions of sequence homology. Adapted from McVey and Lee (2008).

1.3.3. Complement Factor I (FI)

Complement Factor I (FI) is a component of complement that regulates C3 convertase formation (Kavanagh *et al.*, 2008, Fremeaux-Bacchi *et al.*, 2004). It is a serine protease and together with a cofactor, it is able to inactivate C3b and C4b by removing their alpha chains (Kavanagh *et al.*, 2005). This inhibits the formation of C3 and C5 convertases (Caprioli *et al.*, 2006) and in doing so regulates the complement pathway by preventing over activation.

CFI is found on chromosome 4 (Goldberger *et al.*, 1987) and contains 13 exons and encodes the protein FI, an 88kD glycoprotein that is mainly produced by the liver (Vyse *et al.*, 1994). FI is composed of a heavy and light chain joined by a disulphide bond (Roversi *et al.*, 2011). The light chain contains the catalytic serine protease domain (Campbell *et al.*, 1988). The heavy chain does not have catalytic activity, instead was hypothesised to mediate the interaction of FI with cofactors and C3b (Sanchez-Gallego *et al.*, 2012) or be responsible for producing an inactive conformation when FI is circulating in the blood stream (DiScipio, 1992). This modulates the protein activity by 'hiding' the catalytic site, which can then be exposed upon interaction with cofactors (Morgan *et al.*, 2011).

Fremeaux-Bacchi *et al.* (2004) and Kavanagh *et al.* (2005) reported mutations in exons 11 and 13 of *CFI*, which encode the catalytic site of the protein's light chain. Both studies found two mutations that created a premature stop codon, causing truncation of the light chain and thus absence of the serine protease domain. This led to lower FI levels in patients, either due to a reduction in protein secretion or the half-life of the

protein. This consequently led to decreased complement regulation and onset of disease. Other mutations have been reported patients with normal FI levels in the serum, suggesting that they affect function (Caprioli *et al.*, 2006). Nilsson *et al.* (2010) and Kavanagh *et al.* (2008) demonstrated that these mutants have abrogated protease activity and were unable to inactivate C3b and C4b.

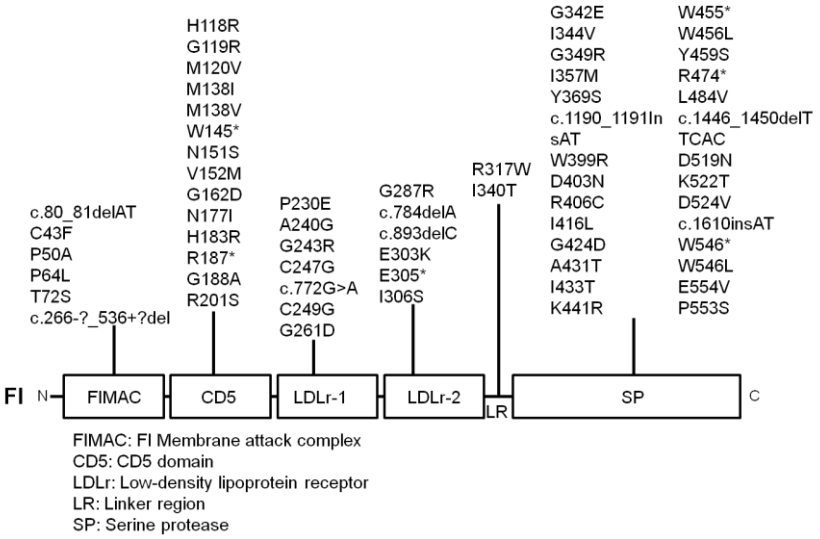

Figure 1-9 FI protein with mutations.

Structure adapted from Sanchez-Gallego et al. (2012). Mutations obtained from aHUS database (Rodriguez et al., 2014) and from Kavanagh et al. (2013).

The frequency of *CFI* mutations is lower than compared with *CFH*. Sellier-Leclerc *et al.* (2007), Kavanagh *et al.* (2005) and Caprioli *et al.* (2006) found that mutations occurred with a frequency of approximately 5% in their respective cohorts. It is attributed to a poor prognosis in patients, with early onset of disease, disease recurrence and in 50% of cases leading to ESRF or death (Caprioli *et al.*, 2006).

1.3.4. Membrane Cofactor Protein (MCP)

Membrane Cofactor Protein (MCP), also known as CD46, is a cell surface protein that is involved in regulating the complement pathway (Liszewski *et al.*, 2000). The extracellular structure of MCP comprises of 4 N-terminal CCP domains, followed by a serine-threonine-proline rich region (STPR) (Lublin *et al.*, 1988). Then there is a transmembrane domain that anchors MCP onto the cell surface and a cytoplasmic tail (Liszewski *et al.*, 1991). It is found as multiple isoforms, depending on alternative splicing of the C-terminal STPR and cytoplasmic tail (Post, 1991). MCP acts as a

cofactor to FI in the cleavage of C3b and C4b, but unlike FH, MCP does not accelerate the decay of the AP convertase (Seya *et al.*, 1986). Studies have identified that CCPs 1 and 2 are important for C4b binding (Liszewski *et al.* 2000) and CCPs 3-4 were involved in C3b and C4b binding (Adams *et al.*, 1991). Impaired ligand binding reduces cofactor activity, but not vice versa (Barilla-LaBarca *et al.*, 2002). MCP is expressed on many surfaces including platelets (Yu *et al.*, 1986) and endothelial cells (McNearney *et al.*, 1989).

MCP is located within the RCA cluster on chromosome 1 (Lublin *et al.*, 1988). *MCP* mutations have been reported in several cohorts with a frequency of approximately 5-15% (Loirat and Fremeaux-Bacchi, 2011, Sellier-Leclerc *et al.*, 2007). *MCP* mutations are less severe with fewer patients developing ESRF. Patients that do develop ESRF are less likely to have disease recurrence after transplantation because the underlying defect in the kidney has been corrected (Caprioli *et al.*, 2006).

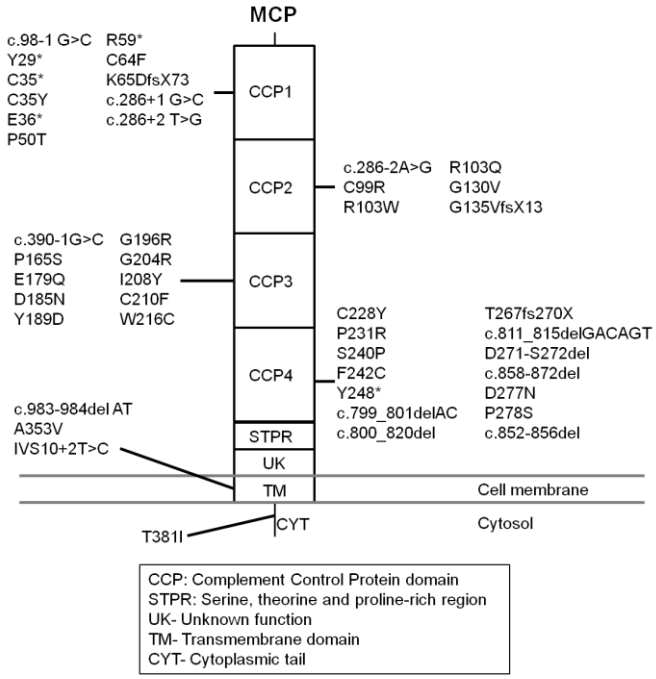

Figure 1-10 MCP protein with mutations.

Protein structure adapted from Lublin et al. (1988). Mutations obtained from aHUS database (Rodriguez et al., 2014) and Kavanagh et al. (2013).

The vast majority of sequence variants identified in *MCP*, approximately 75%, result in the loss of protein expression on the cell surface, whilst the other 25% lead to the formation of a non-functional protein (Kavanagh *et al.*, 2013).

1.3.4.1. MCP risk haplotype

Several Single Nucleotide Polymorphisms (SNPs) were found within the *MCP* gene, termed *MCPggaac* and this haplotype was significantly associated with aHUS (Esparza-Gordillo *et al.*, 2005). These variants are shown in Table 4.

SNP	dbSNP ID	MAF (%)
-652A>G	rs2796267	39.1
-366A>G	rs2796268	36.4
c.989-78G>A	rs1962149	29.3
c.1127+638A	rs859705	31.7
c.*897T>C	rs7144	34.7

Table 4 MCP aHUS risk SNPs.

Minor allele frequency (MAF) is given as a percentage, based on the 1000g database. '' indicated that the substitution occurred 3' of the translation termination codon.*

Esparza-Gordillo *et al.* (2005) found that the frequency of this haplotype was 2 fold higher in aHUS patients than in controls. This was also replicated in French and Newcastle aHUS cohorts (Fremeaux-Bacchi *et al.*, 2005). There were various observations in the significance of the association seen in aHUS patients with and without additional complement mutations, such as in *CFH*, *MCP* or *CFI*. Some cohorts found patients with no additional complement gene mutation had a risk of disease, comparable to that of controls (Esparza-Gordillo *et al.*, 2005), whilst in another cohort these patients had an increased risk of aHUS (Fremeaux-Bacchi *et al.*, 2005). These differences may be the result of small sample sizes. *In vitro* experiments have suggested that this haplotype reduced transcriptional activity by approximately 25% (Esparza-Gordillo *et al.*, 2005). However no change in the surface expression of MCP was seen in human umbilical vein endothelial cells (HUVECs) *in vivo* (Frimat *et al.*, 2012).

1.3.5. Complement Factor B (FB)

Factor B (FB) is one subunit of the AP C3 convertase enzyme, produced by the liver (Maga *et al.*, 2010). *CFB* is located in the Major Histocompatibility Complex (MHC) class III region 6p 21.3 and is composed of 18 exons (Campbell, 1987). It has three CCP domains, one von Willebrand (vWA) type A domain and a serine protease domain (Milder *et al.*, 2007). FB binds to C3b to form a pro-C3 convertase, which is activated when FB is fragmented by FD (Forneris *et al.*, 2010). FD cleaves the linker region between the CCPs and VWA allowing the Ba subunit, containing the CCPs, to be

liberated (Torreira *et al.*, 2009). The active Bb subunit containing the serine protease and vWA domains, remains attached to C3b to form the active C3 convertase (Campbell and Porter, 1983).

CFB mutations are rare, accounting for approximately 1-5% of aHUS cohorts (Maga *et al.*, 2010, Fremeaux-Bacchi *et al.*, 2013). Mutations that occur in this gene lead to enhanced enzyme activity, faster convertase formation or reduced sensitivity to decay. These activating mutations are associated with lower serum C3 levels in patients, due to up regulated complement activity.

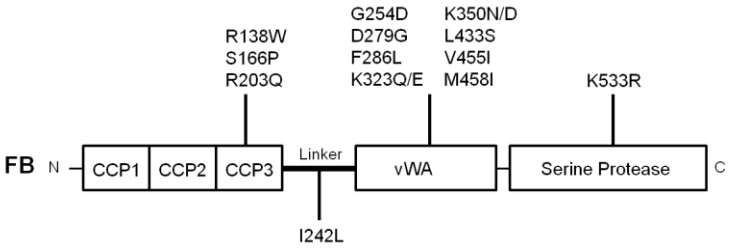

Figure 1-11 FB protein with mutations.

Disease-associated mutations were obtained from the literature and FB protein structure was adapted from Milder et al. (2007).

Two mutations (F286L and K323E) were found in the C3b binding site on the vWA domain and are thought to lead to enhanced C3b:Bb binding and thus faster convertase assembly, with a potentially longer half life (Goicoechea de Jorge *et al.*, 2007). In addition they were less sensitive to decay in the presence of FH and DAF. Another aHUS case reported a K533R located in the serine-protease domain (Tawadrous *et al.*, 2010). This was shown to enhance convertase activity, breaking down even more C3 (Tawadrous *et al.*, 2010). Hourcade *et al.* (2002) mapped DAF binding sites on FB by generating mutant FB proteins. They found that there is an area on the vWA domain that is important in DAF interaction, which would explain why these mutations are more resistant to decay in the presence of DAF.

1.3.6. Complement component 3 (C3)

Complement component 3 (C3) is a serum protein produced by the liver and plays a critical role in the complement system (Fearon *et al.*, 1976, Maga *et al.*, 2010). It is 186kD in size until it is cleaved by C3 convertases, when it forms a 9kD (C3a) and 177kD (C3b) fragments (Janssen *et al.*, 2006). C3b can then interact with FB to form

the AP pro-C3 convertase, producing a positive feedback loop. To prevent deregulation, C3b is rapidly degraded by FI along with a cofactor, such as FH, CR1 and MCP (Lambris *et al.*, 1988).

C3 is positioned on chromosome 19p13.3 (Shaw *et al.*, 1986). It encodes a protein consisting of an α (amino acids 650-1641) and β (amino acids 1-645) chain (Janssen *et al.*, 2005), linked by disulphide bridges (Thomas *et al.*, 1982). Cleavage of C3 into C3b results in the removal of the ANA domain, which forms the C3a fragment (Janssen *et al.*, 2005). The side chains of amino acids C988 and Q991 form a thioester bond, which maintains the conformation of the protein (Thomas *et al.*, 1982). Hydrolysis of this thioester bond produces C3(H_2O), which is similar in conformation to C3b and can also bind to FB (Isaac and Isenman, 1992). Residues 727-767 of the N-terminus of cleaved α -chain (α'NT) domain have been demonstrated to be involved in C3 binding to FB, FH and CR1 (Oran and Isenman, 1999).

Similar to *CFB* mutations in *C3* are activating, causing enhanced C3 convertase activity and faster convertase formation. Mutations in *C3* account for an estimated 4-8% of aHUS cases (Maga *et al.*, 2010, Fremeaux-Bacchi *et al.*, 2013, Noris *et al.*, 2010), summarised in Figure 1-12.

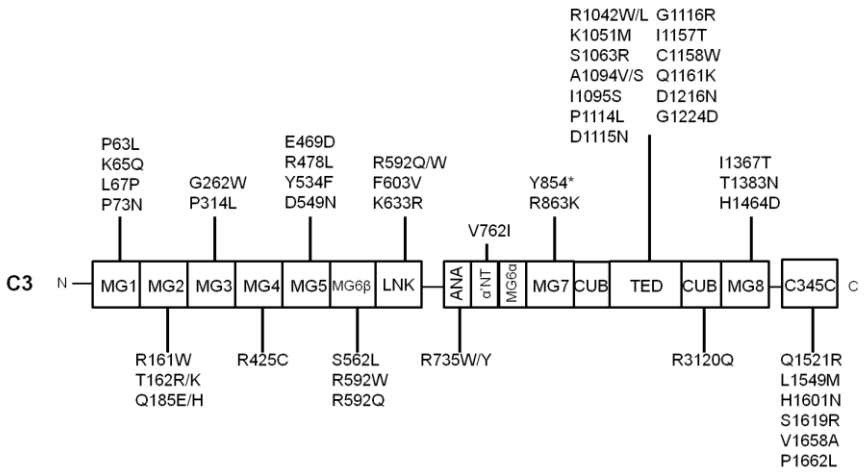

Figure 1-12 C3 protein with mutations.

MG= macroglobulin domain, LNK= linker domain, ANA= anaphylatoxin domain, α'NT= N-terminus of cleaved α-chain, CUB= complement C1r/C1s, Uegf, Bmp1 domain, TED= thioester-containing domain. Mutations were compiled from the literature (Schramm et al., 2015, Rodriguez et al., 2014, Kavanagh et al., 2013). Protein structure adapted from Janssen et al. (2005).

Fremeaux-Bacchi *et al.* (2008) found 5 heterozygous mutations (R570Q, R570W, A1072V, D1093N and Q1139K) that affected the interaction of C3 with MCP. When compared to wild type C3, these mutants had reduced binding with MCP, but normal FB binding (Fremeaux-Bacchi *et al.*, 2008). As a result mutant AP convertases were formed normally but due to decreased cofactor activity of MCP, had longer half lives. This would result in more C3b generation, leading to increased complement activity.

1.3.7. Thrombomodulin (THBD)

Thrombomodulin (THBD) is a cell surface bound protein that can inhibit the activity of thrombin and enhance the activity of protein C, reducing coagulation (Anastasiou *et al.*, 2012). *THBD* is located on chromosome 20p11.2 (Maglott *et al.*, 1996) and found to lack introns (Jackman *et al.*, 1987). THBD is composed of an N-terminal Lectin-like domain, 6 tandemly repeated epidermal growth factor-like (EGF) domains, serine/threonine rich segment, transmembrane region and a cytoplasmic tail (Wen *et al.*, 1987). EGF4, 5 and 6 are critical for protein C cofactor activity (Gale and Griffin, 2004).

Figure 1-13 THBD protein with mutations.

L= Lectin-like domain, EGF= epidermal growth factor- like domain, STR= serine/threonine rich segment, TM= transmembrane region and CYT= cytoplasmic tail. Adapted from Tsiang et al. (1992). Mutations obtained from the literature.

Autosomal dominant *THBD* sequence variants have been previously reported in aHUS patients by Delvaeye *et al.* (2009) and Maga *et al.* (2010). The frequency of a known SNP (A473V) was not significantly different between the aHUS and control cohort (Delvaeye *et al.*, 2009). Functional analysis of THBD (Delvaeye *et al.*, 2009) demonstrated that THBD acted as a cofactor to FI in the presence of C4BP and FH, leading to increased inactivation of C3b. They established that mutant THBD was less able to convert C3b into iC3b, despite increased binding to FH and C3.

1.3.8. MMACHC

Cobalamin C disease is a disorder of cobalamin (Cbl), or vitamin B_{12}, metabolism (Sharma *et al.*, 2007, Martinelli *et al.*, 2011). It can occur as a result of autosomal recessive sequence variants in methylmalonic aciduria *cblC* type with homocystinuria (*MMACHC*) gene (Lerner-Ellis *et al.*, 2006). MMACHC is involved in reducing the cobalt atom of Cbl and transporting Cbl within the cell (Green, 2010). Loss of MMACHC leads to the accumulation of homocysteine, increasing the risk of TMA (Cattaneo, 1999, Morel *et al.*, 2006). Generally this disease occurs in infancy, although there has been a reported case of an adult developing cobalamin C- associated HUS, who failed to respond to eculizumab (Cornec-Le Gall *et al.*, 2014). It is diagnosed by high levels of homocysteine in the blood and methylmalonic aciduria (Loirat *et al.*, 2012).

1.4. Two hit hypothesis

It is now known that aHUS is caused by deregulation of the complement pathway, due to mutations in genes encoding complement proteins (Kavanagh *et al.*, 2013, Fremeaux-Bacchi *et al.*, 2013, Maga *et al.*, 2010, Noris *et al.*, 2010). However it has been

proposed that to acquire aHUS, an initiating environmental event is required to activate the complement system (Caprioli *et al.*, 2006). The complement system then becomes deregulated as a result of a genetic defect, leading to the onset of disease.

One study found that 70% of aHUS patients with a *CFH* mutation had an infection prior to disease onset, whilst 4% had either been pregnant or taking certain prescribed drugs (Abarrategui-Garrido *et al.*, 2008). In the case of patients with *CFI* mutations, they established that 40% individuals were pregnant and 40% had an infection prior to aHUS onset (Abarrategui-Garrido *et al.*, 2008). This is also seen in French and Italian paediatric cohorts, where aHUS was preceded by an infection in 63% and 85% of cases, respectively (Loirat *et al.*, 2008). This could explain how within a family, individuals can acquire the disease at any point within their lives, with varying severity. Fakhouri *et al.* (2010) identified pregnancy as a trigger for aHUS, with 20% of the females having disease onset during pregnancy.

Roumenina *et al.* (2012) also noted that 70% of patients with a sequence variant in *C3* had aHUS after a trigger event and that penetrance was incomplete. In affected individuals the presence of additional risk haplotypes such as MCP_{ggaac} and *CFH*-H3 were higher than in healthy carriers. This suggested that several factors may be required to lead to disease onset.

1.5. Treatment

Eculizumab is a humanized monoclonal antibody against C5, a component of MAC (Rother *et al.*, 2007). This antibody binds to C5 and thus prevents the formation of the MAC, reducing the activation of prothrombotic and proinflammatory pathways (Meri, 2013). Failure to respond to eculizumab may suggest that complement is not involved in disease pathogenesis or that treatment was initiated too late to be effective (Wong *et al.*, 2013). The exception to this is that there is a known sequence variant in *C5*, p.R885H, which has been shown to reduce drug efficacy (Nishimura *et al.*, 2014). Therefore all patients who will receive this treatment by the national aHUS service are routinely tested for this sequence variant (Sheerin *et al.*, 2015).

1.6. Whole Exome Sequencing (WES)

Next Generation Sequencing (NGS) is becoming an increasingly common method of genetic analysis for Mendelian diseases (Botstein and Risch, 2003), particularly Whole Exome Sequencing (WES). This was first used to successfully identify the genetic cause of Freeman-Sheldon syndrome (Ng *et al.*, 2009). It has since been used to discover novel genetic causes of many diseases. The method of exome capture was first described by Hodges *et al.* (2007), which when combined with next generation sequencing, produced a very rapid approach to sequencing large areas of DNA.

Traditional methods for candidate gene identification were to use Sanger sequencing (Ku *et al.*, 2012), linkage analysis (Kerem *et al.*, 1989, Riordan *et al.*, 1989), homozygosity mapping (Lander and Botstein, 1987) or Genome-wide association studies (GWAS) (McCarthy *et al.*, 2008). However these techniques have limitations which are described in section 5.1.

1.6.1. Advantages

WES utilizes the sequence data from the exonic regions of DNA only, approximately 1% of the entire genome. Despite this, the exome is estimated to contain 85% of disease causing mutations (Choi *et al.*, 2009). Therefore it has the benefit of being more cost effective than targeted resequencing and much faster than whole genome sequencing (WGS).

WES is an unbiased approach to finding candidate variants and allows for the screening of genes that may not be within a pathway of interest (Boycott *et al.*, 2013). This is useful for screening genes that have associations with alternative diseases, which may be important if patients have atypical presentation of disease or have been misdiagnosed (Bamshad *et al.*, 2011). WES has also been demonstrated to be able to detect mosaicisms, where the ratio of variant alleles compared to reference alleles is lower and tissue-specific (Riviere *et al.*, 2012).

1.6.2. Disadvantages

There are several potential disadvantages to WES. Firstly it excludes intronic data, which comprises approximately 99% of the genome. This could lead to potentially important mutations being missed. There are also some technical issues such as it

poorly sequences areas that are CG-rich or highly repetitive (Aird *et al.*, 2011, Kozarewa *et al.*, 2009) and generally tends to favour the wild type allele (Meynert *et al.*, 2014).

Secondly WES generates a huge number of variants and it can be difficult to determine which are pathogenic and which are bystanders (Bamshad *et al.*, 2011). Examining variants that are not in the general population, or are rare, is one method to select candidates. However it is thought that each person can have around 30 de novo variants, which can make this strategy more complicated (Marian, 2012). Variants of unknown significance (VUSs) are another major problem of next generation sequencing, particularly if they occur in a gene with limited functional information (Goldstein *et al.*, 2013). The use of prediction software can help to prioritise sequence variants, however they are only a prediction and can often be non-concordant (Tennessen *et al.*, 2012).

Finally this analysis assumes that this disease is a monogenic disorder and that the occurrence of one causative gene defect leads to disease. However it may be that there are additional genetic variations that contribute to disease or additional environmental factors, which have been previously reported in aHUS (Caprioli *et al.*, 2006).

1.7. Aim of project

At the time the project commenced (2010) approximately 45% of aHUS cases in the Newcastle familial cohort had a known genetic cause of disease, shown in Figure 1-14. This is comparable to what has been described in other studies (Noris *et al.*, 2010, Maga *et al.*, 2010, Fremeaux-Bacchi *et al.*, 2013). Currently 7 genes have been associated with aHUS (*CFH*, *MCP*, *CFI*, *CFB*, *C3*, *THBD* and *MMACHC*). The largest proportion of patients have mutations in *CFH*, followed by *C3* and *MCP*, with *CFI* mutations were seen less frequently. No familial cases have been identified with genetic abnormalities in *CFB*, *THBD* or *MMACHC*. However there have been reports of sequence variants occurring in *CFB* in the sporadic cohort.

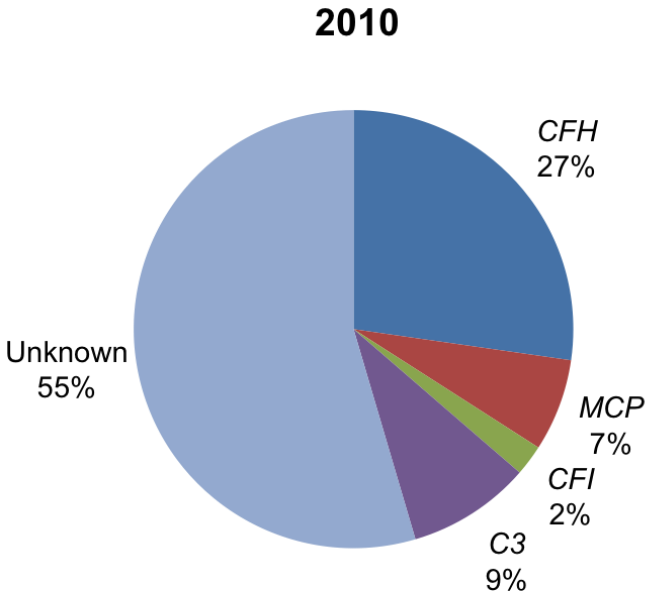

Figure 1-14 Percentage of gene mutations in 2010.

Percentage of gene mutations observed in the Newcastle familial aHUS cohort in 2010, before the start of this project.

The aim of this project is to elucidate the genetic cause of disease in familial aHUS cases, in whom no genetic cause had previously been identified. Initially CNVs occurring in known aHUS-associated genes were identified using Multiplex Ligation-dependent Probe Amplification (MLPA) and Western Blotting, with complement assays used to establish their functional significance. Subsequently WES was used to identify candidate genes using a filtering strategy based on different inheritance patterns, minor allele frequency (MAF), *in silico* analysis of functional significance and conservation and known phenotypic information.

Chapter 2: Methods

2.1. Patient selection

The study was approved by Newcastle and North Tyneside 1 Research Ethics Committee and informed consent was obtained in accordance with the Declaration of Helsinki. Patients in Newcastle aHUS cohort were chosen if there was a family history of disease. It should be noted that some families were referred from outside of the UK.

2.2. DNA and RNA extraction

Performed by the Northern Genetics Service. DNA was isolated from peripheral blood leukocytes using QIAamp DNA mini kit (QIAGEN). RNA was extracted from peripheral blood leukocytes using RNAeasy mini kit (QIAGEN).

2.3. DNA quantification

DNA was quantified using a NanoDrop 8000 (Thermo Scientific) at 260nm.

2.4. Serum preparation

Performed by the NHS. Serum was prepared using a standard method (Whaley and North, 1997). Blood was allowed to clot by incubating at 37°C for 30 minutes. This was then chilled on ice for 1 hour and the serum separated by centrifugation at 2000xg for 5 minutes at 4°C. Serum was removed and stored at -70°C until use.

2.5. PCR protocols

2.5.1. CFH/CFHR3 break point analysis

To confirm the position of the breakpoint that caused the formation of the *CFH/CFHR3* hybrid gene, genomic DNA was amplified using a forward primer specific for *CFH* in exon 20 (GTAACTGTTATCAGTTGATTTGC) and a reverse primer in *CFH* 3'UTR (ACGGATTGCATGTATAAGTG), shown in Figure 3-4. This was performed by the Northern Genetics Service.

2.5.2. CFH/CFHR3 cDNA amplification

Complementary DNA (cDNA) was synthesised by the Northern Genetics Service with SuperScript III First-Strand Synthesis System (Invitrogen) using random hexamers and the extracted patient Ribonucleic acid (RNA) as a template. cDNA was used as a template in a polymerase chain reaction (PCR) reaction with a forward primer in *CFH* exon 20 (TGGATGGAGCCAGTAATGTAAC ATGCAT) and a reverse primer in *CFHR3* exon 2 (GAAATAGACCTCCATGTTTA ATGTCTG).

2.5.3. Primer design

Genomic sequences for genes were downloaded from UCSC Genome Browser (Kent *et al.*, 2002, Karolchik *et al.*, 2014). Sequences of approximately 18-20 nucleotides were then selected either side of the region of interest. The reverse primer was reverse complemented. These sequences were then run through the UCSC *in silico* PCR software , which determined the primer specificity. All primers were tagged with a 17bp sequence that is specific for sequencing primers. This allows the PCR products to be sequenced. The sequences for the forward and reverse primers were 'GTAGCGCGACGGCCAGT' and 'CAGGGCGCAGCGATGAC' respectively.

2.5.4. Polymerase chain reaction (PCR)

PCR was carried out to generate enough DNA transcripts for subsequent Sanger sequencing. 50-250ng/µL of genomic DNA was amplified by PCR comprising of 10µL Immomix Red (Bioline, BIO-25021), 1µL forward primer (10µM), 1µL reverse primer (10µM) and sufficient H_2O to make a final volume of 20µL. PCR was then carried out on a Bio-Rad T100 Thermal Cycler (186-1096).

Cycle conditions:

1. 95 °C for 10 minutes
2. 94 °C for 1 minute
3. Annealing temperature* for 1 minute
4. 72 °C for 2 minutes. Return to step 2 and repeat 34 times.
5. 72 °C for 10 minutes
6. 4°C forever.

*For annealing temperatures and primer sequences, see tables Table 33 and Table 34 in Appendix B.

2.5.5. Agarose Gel Electrophoresis

PCR samples were run on an agarose gel prepared from 1% (w/v) Low EEO Agarose (NBS Biologicals Ltd, NBS-AG500), 1x TAE (40 mM Tris, 20 mM Acetic acid and 1 mM EDTA) and 0.0001% (v/v) GelRed (Biotium, 41003). 5µL of DNA ladder (MBI Fermentas, SM0383, New England Biolabs, N3231s or N3232s) and 5µL of PCR product were loaded onto the agarose gel. This was run in 1x TAE, at 70 volts for 30 minutes, using PowerPac Basics (Bio-Rad. 300V, 400mA, 75W). DNA bands were viewed using GelDoc-it 310 Imager (UVP).

2.5.6. Sanger sequencing

If PCR products were generated, samples were purified using QIAquick PCR purification Kit (QIAGEN, 28104), ready for sequencing. If PCR was performed in 48 or 96 well plates then purification was performed by GATC Biotech. Typically samples were eluted with 30µL of H_2O and sent to GATC Biotech for Sanger sequencing. Sequencing was performed using BigDye Terminator v3.1 (Life Technologies) and analysed on the Applied Biosystems 3730xl DNA Analyzer (Life Technologies). Sequences were then viewed using Sequencher (Gene Codes Corporation, Version 5.0).

2.5.7. Sequence variant nomenclature

Sequence variants were named using the recommendations given by the Human Genome Variation Society (den Dunnen and Antonarakis, 2000).

2.5.8. Multiplex ligation-dependent probe amplification (MLPA)

2.5.8.1. Background

Multiplex ligation-dependent probe amplification (MLPA) is a PCR-based method (Schouten *et al.*, 2002, Eldering *et al.*, 2003), which can identify insertions or deletion in a region of interest. It has the benefit of detecting very small base changes and being a rapid, cost effective technique (Stuppia *et al.*, 2012). However in some instances, MLPA will not pick up gene rearrangements. For example it is difficult to detect CNVs

in genes that have complete sequence homology. In these instances abnormalities would only be found by looking at the protein itself, via western blotting.

MLPA probes are formed of two oligonucleotides that to bind to the region of interest via a complimentary sequence, known as the hybridisation sequence. The oligonucleotides are then ligated by a ligase enzyme, which allows for PCR to take place during the subsequent amplification steps. If ligation does not take place, then amplification does not occur, the aim of which is to minimise amplification of probes that bind non-specifically. Each MLPA probe contains a stuffer sequence that gives the probe a specific length, allowing them to be separated during electrophoresis. The amount of MLPA probes produced after amplification is proportional to the number of copies of the target sequence in the sample. This can then be compared to a control sample to determine if there are normal or abnormal numbers of gene copies. Where possible, at least one individual per family was tested for CNVs in *CFH*, *CFHRs*, *CFI* and *MCP*.

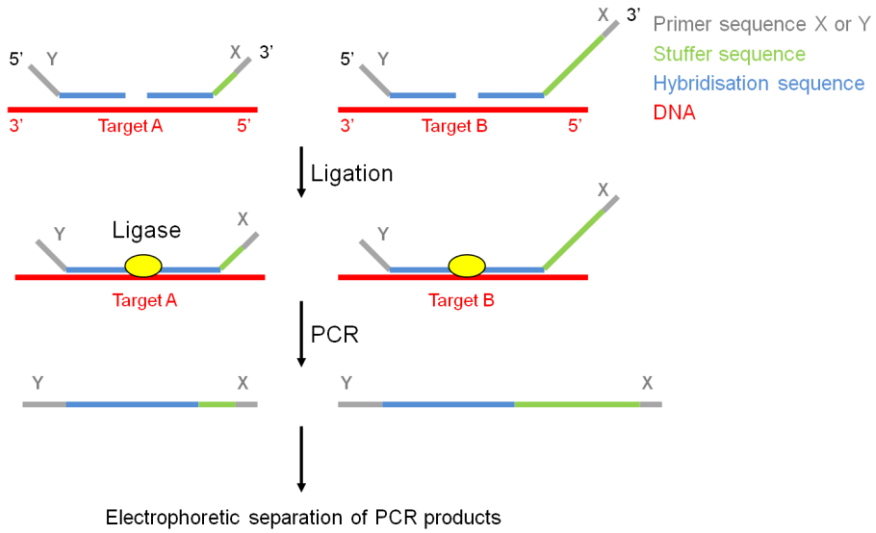

Figure 2-1 Schematic diagram showing MLPA methodology.

Each MLPA probe is composed of two oligonucleotides. They anneal to the target position on the DNA and then ligated using a ligase enzyme. PCR is then carried out to amplify the MLPA probes, which are then separated by electrophoresis. Grey indicates the standard probe sequence used specifically for either left or right probe part. Green shows the stuffer sequence. Blue shows the hybridisation sequence. Red indicates the DNA target sequence. Figure adapted from Schouten et al. (2002).

2.5.8.2. CFH and CFHRs MLPA probes

Patients with a potential diagnosis of aHUS are screened routinely by the Northern Genetics Service for CNVs in the RCA cluster. They screen *CFH*, *CFHR2*, *CFHR3* and *CFHR5* using the SALSA MLPA probemix P236-A1 ARMD (MRC Holland), probe sequences can be found in Appendix D, Table 36. An additional high density MLPA for *CFH* and *CFHR5*, designed in house by Geisilaine Soares dos Reis Araujo. This was used to provide additional coverage, probe sequences can be found in Table 35 of Appendix C. Figure 2-2 shows the position of the MLPA probes in *CFH* and *CFHRs*. The probes used for *CFH* exons 21 and 22, designed in house, cross-react with *CFHR1* and therefore were not included in MLPA analysis.

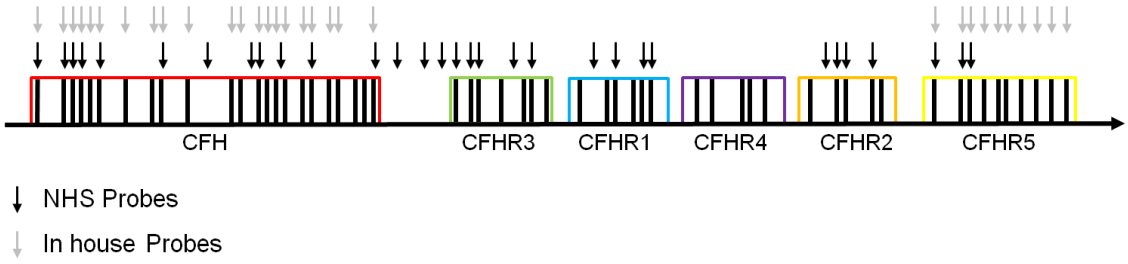

Figure 2-2 MLPA probe positions for CFH and CFHRs.

Diagram shows the position of CFH and CFHRs in the RCA cluster. Each black line represents an exon and genes are highlighted by different colours. Black arrows indicate probes used by the Northern Genetics Service for routine testing. Grey arrows show the position of MLPA probes, designed in house for CFH and CFHR5.

2.5.8.3. CFI and MCP MLPA probes

The Northern Genetics Service also used a probe kit used to look for CNVs in *CFI* and *MCP* was SALSA MLPA probemix P296-A2 (MRC Holland). Table 37 in Appendix D shows the MLPA probe hybridization sequences.

2.5.8.4. MLPA reaction

Reactions were performed with a positive control, containing either a deletion or duplication and a negative control, containing the normal number of gene copies. 100-200ng of genomic DNA was used per reaction. MLPA was carried out according to the manufacturer's instructions (MRC-Holland, 2013), using SALSA MLPA P200 Human DNA reference-1 probemix (MRC-Holland, P200-100R), SALSA MLPA EK1 reagent kit (MRC-Holland, EK1-FAM) and MLPA probes for *CFH or CFHR5*, designed in

house. The amplification products were then run on an Applied Biosystems 3130 Genetic Analyzer (Life Technologies). Data was analyzed using GeneMarker software, version 4.2 (SoftGenetics LLC). Values between 0.8 and 1.2 were considered to be within the normal range. These experiments were carried by the Northern Genetics Service, Geisilaine Soares dos Reis Araujo and myself.

2.5.8.5. MLPA results

The result of the in house *CFH* and *CFHR5* and diagnostic *CFI*/*CD46* MLPA were plotted in Excel. The x axis shows the gene and exon tested and the y axis shows the copy number, which is calculated by comparing the peak heights of a control and patient sample. Results are shown in the Supplementary data. Graphs could not be drawn for the SALSA MLPA probemix P236-A1, as the original data was not available.

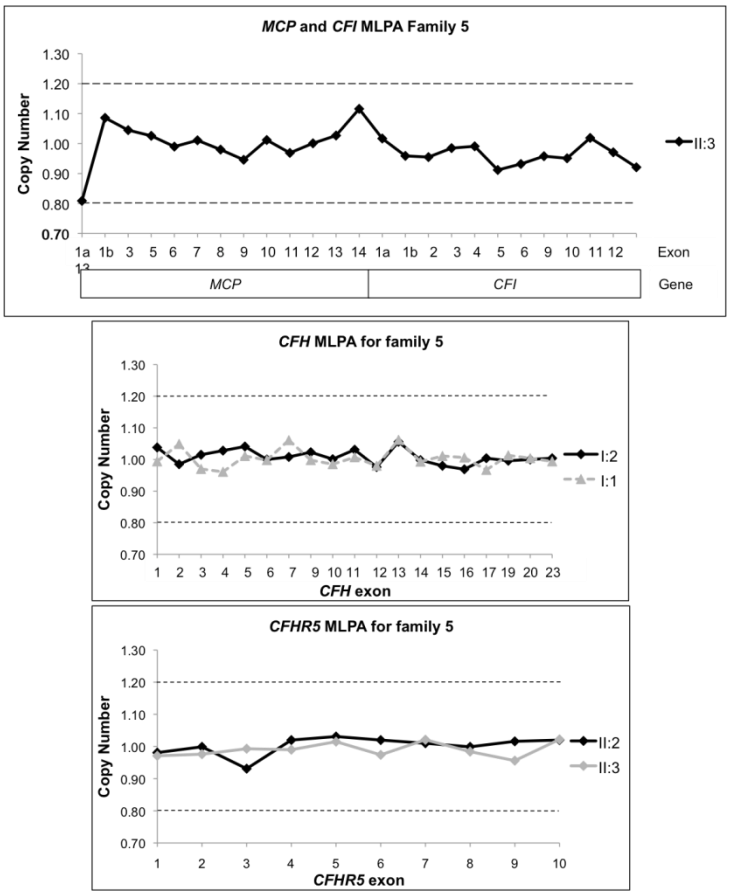

Figure 2-3 Examples of MLPA data

Examples of the MLPA data obtained using the MCP/ CFI probe kit and the in house CFH and CFHR5 probe kits.

2.5.9. Whole Genome Amplification

Patients with small quantities of DNA available had Whole genome amplification (WGA) carried out. This was done using either the GenomePlex Complete WGA kit (Sigma- Aldrich, WGA2- 50RXN), Repli-G mini kit (QIAGEN, 150023) or Repli-G FFPE mini kit (QIAGEN, 150243) for formalin fixed paraffin embedded (FFPE) tissue. See Appendix H, Table 42 to see the method used on specific samples. WGA samples were then purified using GenElute PCR Clean-up kit (Sigma- Aldrich, NA1020), to remove any remaining primers, nucleotides or other contaminants that may inhibit downstream experiments.

2.6. Whole exome sequencing

2.6.1. Background

Whole exome sequencing (WES) is a method used to sequence the coding region of the genome. To isolate the exonic regions of the DNA, specialist kits are used. In general it is based on the use of synthetic oligonucleotides, to pull out DNA fragments containing corresponding exon-specific sequences (Hodges *et al.*, 2007). The captured DNA is then fixed to the surface of a flow cell, where it is amplified prior to sequencing (Metzker, 2010). This produces numerous clusters, containing identical copies of DNA, increasing the intensity of the fluorescence and thus allowing for more accurate detection. The sequencing chemistry used in this project was based on Sequencing by Synthesis (SBS) technology developed by Ju *et al.* (2006), where nucleotides are detected as they are incorporated into the newly synthesised strand. This can be done due to the presence of a fluorophore attached to the dNTP (deoxynucleotide triphosphates), causing sequence termination (Guo *et al.*, 2008). Cleavage of the flurophore allows the addition of the next dNTP, which is repeated until the desired read length is achieved (Ju *et al.*, 2006). Each dNTP has a specific fluorophore, which allows the detector to differentiate between them according to their emission spectra (Metzker, 2010).

There are some problems with each step. Exome capture has been shown to have some selection bias according to exon length (Hodges *et al.*, 2007) and GC content (Clark *et al.*, 2011). Amplification of DNA on the flow cell may introduce sequencing errors and if clusters get two big or are positioned too close together, then the fluorescent signals may interfere (Rizzo and Buck, 2012, Mamanova *et al.*, 2010). Finally sequencing may be problematic due to interference from unbound dNTPs, overlapping emission wavelengths from fluorophores or the introduction of errors as the read length increases (Fuller *et al.*, 2009).

2.6.2. Sequencing

DNA was sent to sequencing providers AROS AB and GATC Biotech. Three exome enrichment kits were used, TruSeq exome enrichment kit (Illumina), Nextera Rapid Capture Exome kit 37Mb (Illumina) and SureSelect^{XT} Human all exon V5 (Agilent). To see which enrichment kit was used on which sample, see Table 42 in Appendix H

DNA library formation required 1µg gDNA, when using Illumina kits and 3µg gDNA for the Agilent kit. Library preparation involved DNA fragmentation, fragment ligation to adaptors and amplification, using proprietary methods of AROS AB or GATC Biotech. DNA libraries were then amplified in order to add index tags and then enriched for exonic DNA. After enrichment libraries were then denatured and added to the flow cell. A final amplification step is required to produce clusters of identical DNA strands on the surface of the flow cell. Finally sequencing was carried out on a HiSeq 2000 (Illumina). A schematic diagram showing these steps is shown in Figure 2-4.

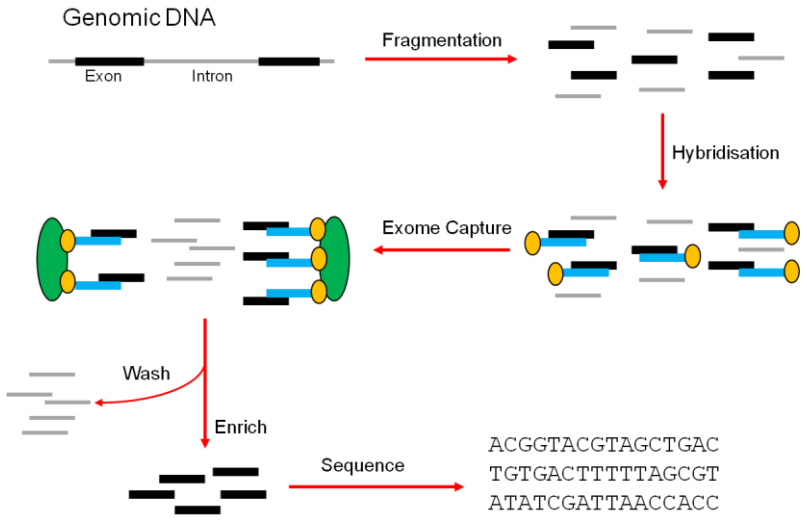

Figure 2-4 Diagram showing WES preparation steps

Genomic DNA (exons are in black and introns are in grey) is fragmented, then during a hybridisation step the exome is captured by capture probes (shown in blue), which have sequences complimentary to the exome. These probes are biotinylated (orange circles), which allows them to be captured by streptavidin beads (shown in green). The unbound sequences are then removed during wash steps, leaving the final enriched exome library. This is then eluted and added to the flow cell ready for cluster amplification and sequencing steps. Adapted from Ku et al. (2012).

2.6.3. Bioinformatic pipeline

The sequencing reads were analyzed using the following workflow, designed and performed by Yaobo Xu and Mauro Santibanez-Koref, to generate a list of variants per family, demonstrated in Figure 2-5. The first step was to perform quality control (QC) on the sequencing reads, performed using FastQC (Andrews). Duplicate reads that arise during library amplification steps, were removed with FastUniq (Xu *et al.*, 2012). The remaining reads were mapped to the human reference genome (GRCh37) with BWA (Li and Durbin, 2010). The alignments were refined with tools of the GATK suite

(McKenna *et al.*, 2010). Variants were called according to GATK Best Practice recommendations (DePristo *et al.*, 2011, Van der Auwera *et al.*, 2002), including recalibration. Freebayes was also used to call variants from the same set of samples (Garrison and Marth, 2012). The variants called by Freebayes with total coverage ≥5, minor allele coverage ≥5 and variant call quality ≥20, were added to those identified by GATK. Annovar was used for annotations and prediction of functional consequences (Wang *et al.*, 2010).

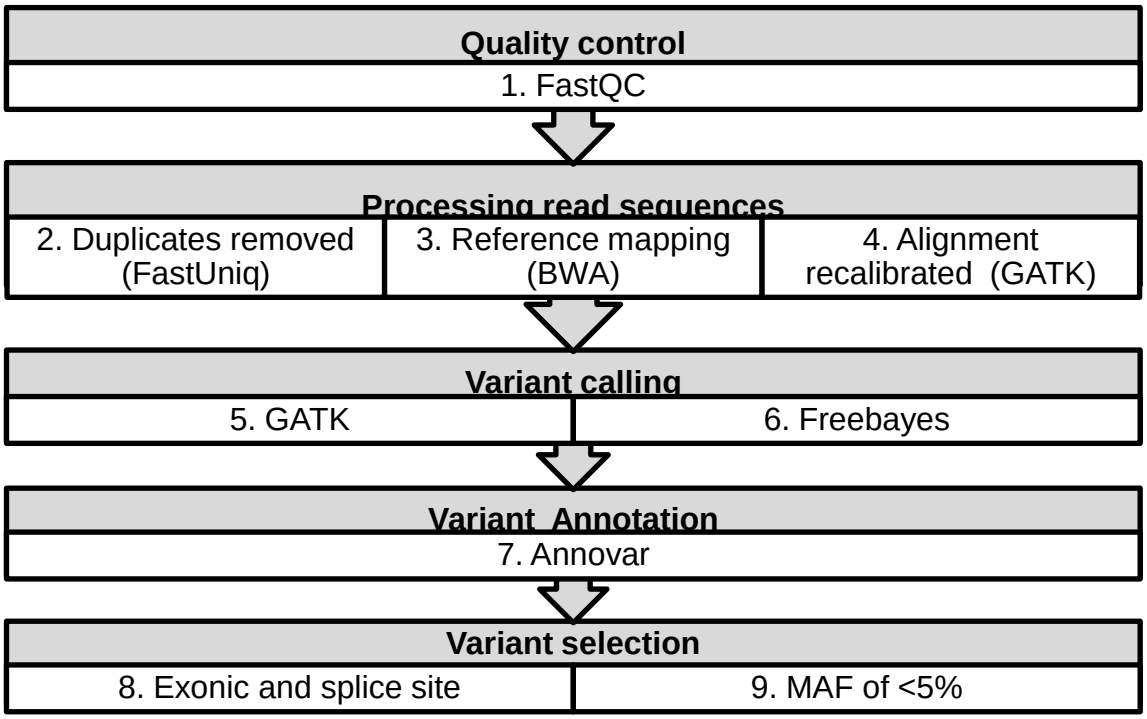

Figure 2-5 Schematic diagram showing bioinformatic workflow.

Variants that were not classed as exonic or in the 2bp region adjacent to exons, according to the Ensembl database (Cunningham *et al.*, 2015), were removed. Then variants with MAF >5% in the 1000 Genomes (1000g) project (2012 Feb release) (Abecasis *et al.*, 2012) and NHLBI GO Exome Sequencing Project (ESP6500) (Furukawa *et al.*, 2011), were excluded. Table 5 gives additional information on the individual software used in the bioinformatic pipeline.

Name	Function	Reference
FastQC	Performs a series of quality checks on raw sequence data. It examines the quality score per base/ sequence, per base sequence content, GC content per base/ sequence and the sequence duplication level.	(Andrews)
FastUniq	Identifies and removes duplicate reads, by comparing sequences within paired reads.	(Xu *et al.*, 2012)
BWA	Maps reads to a reference genome.	(Li and Durbin, 2010)
GATK	Performs local de novo assembly and realigns INDELs. It also calls Insertions/deletions (INDELs) and Single Nucleotide Variant (SNVs), comparing them with known variants of good quality (from 1000g), to see if called variants are true.	(McKenna *et al.*, 2010)
FreeBayes	This detects variants in alignment data by haplotype-based methods, using a Bayesian statistical framework.	(Garrison and Marth, 2012)
Annovar	A tool that utilises data from various databases to annotate SNVs.	(Wang *et al.*, 2010)

Table 5 Description of tools used in WES bioinformatic workflow.

2.6.4. Coverage

To see if genes of interest were covered by WES, .bam files were loaded onto Integrative Genomics Viewer (IGV) (Thorvaldsdóttir *et al.*, 2013, Robinson *et al.*, 2011). This allowed us to see the number of sequencing reads there were at a given position. A position with coverage of more than 20 reads was deemed to be well covered. However all candidate variants were confirmed by Sanger sequencing.

2.6.5. Ingenuity variant analysis (IVA)

Another method used to look for disease-causing variants was Ingenuity Variant Analysis (IVA) (QIAGEN) . It is a web-based platform that utilises evidence-based information, annotation algorithms and other computational software, to allow users to manipulate WES data. It was used here as a 'proof-of-principle' tool, using their 'Path-to-phenotype' data. This is a repository of curated publications that give evidence to support a mechanism of disease.

2.6.6. In silico analysis

2.6.6.1. Impact of mutations on proteins

It can be difficult to differentiate sequence variants that would affect protein function and those that would be tolerated. Several prediction tools were used to help identify

variants that were likely to be pathogenic from those that were not. The prediction score thresholds used were based on what had been previously described in the literature (Li *et al.*, 2014, Dong *et al.*, 2015).

PolyPhen-2 is a server that can predict the effects that a mutation would have on a protein (Adzhubei *et al.*, 2010). It does this by searching for similar proteins and aligning the sequences, looking for any variations. A score is allocated to sequence variants depending on the pattern of substitutions, the amino acid changed that has occurred and possible structural features that may be affected. In doing so the software can predict whether a mutation would be damaging or benign . The model was trained on two datasets, giving two scores. The first data set HDIV contains known disease-causing alleles classed as deleterious and variants that are different between human and other mammals, classed as benign. The second dataset HVAR contains known disease-causing alleles classed as deleterious and common variants (MAF>1%) that are classed as benign. Scores of $\geq$0.453 or $\geq$0.447 in HDIV and HVAR respectively, were labelled damaging. Scores below these thresholds were considered to be benign. Scores cannot be calculated for nonsense mutations, splice site changes or INDELs.

Mutation Taster is an online program that can be used to predict the outcome of sequence variants on protein structure and function (Schwarz *et al.*, 2010, Schwarz *et al.*, 2014). It combines information from several databases (Swiss Prot, dbSNP, Ensembl, HapMap) and prediction programs to annotate variant data. It can predict if changes are within a conserved region and calculates allele frequencies . A score of $\geq$0.5 was classed as deleterious. Values below this threshold were classed as neutral and if it was also seen in HapMap database or in more than 4 cases, then reported as a polymorphism.

Mutation Assessor is a server that predicts the impact of an amino acid substitution on a protein, basing this on whether the position is evolutionary conserved or not (Reva *et al.*, 2011). It cannot be used to calculate scores for nonsense mutations, splice site changes or INDELs. Scores >0.785 were classed as high impact and therefore deleterious. Remaining variants were classed as neutral (scores $\leq$0.55), low impact ($0.55 < \chi \leq 0.645$) or medium impact ($0.645 < \chi \leq 0.785$).

Functional Analysis through Hidden Markov Models (FATHMM) is an online platform that can determine if a missense variant will have an effect on a protein (Shihab *et al.*,

2014, Shihab *et al.*, 2013b). It does this by aligning protein sequences of homologous proteins and calculating conservation scores, combined with the tolerance of the protein to mutations (Shihab *et al.*, 2013a). It cannot be used to calculate scores for nonsense mutations, splice site changes or INDELs. Scores of ≥0.453 were predicted to be deleterious. Variants with scores below this level were classed as 'Tolerated'.

RadialSVM is an ensemble-based approach that can combine multiple scoring methods, such as those that determine sequence conservation or effect on protein function (Dong *et al.*, 2015). It cannot be used to calculate scores for nonsense mutations, splice site changes or INDELs. A score of ≥0.5 was classed as deleterious. Variants with scores below this level were classed as 'Tolerated'.

2.6.6.2. Sequence conservation

If a sequence is considered to be evolutionary conserved, then it suggests that it performs a critical function. If mutations occur here, then they are predicted to be significant and potentially disease causing.

PhyloP is a program that examines the conservation of an area where a mutation has occurred. Conserved sequences are areas where evolutionary changes are very slow or slower than "Neutral drift" (Pollard *et al.*, 2010). These areas are often very similar across several species and may indicate that an area is functionally important. The program analyses conservation by aligning 36 sequences of a variety of mammals and detecting differences between them (Pollard *et al.*, 2010), calculating an estimated neutral evolution rate. It then compares the likelihood of observing substitutions at a neutral evolutionary rate. It cannot calculate scores for INDELs. A score >0.5 was considered conserved.

GERP++ is a tool that identifies genomic regions that are evolutionary conserved (Davydov *et al.*, 2010). It aligns sequences from multiple species and determines the neutral evolution rate. It then calculates the number of nucleotide substitutions that would be expected (at a neutral rate of evolution), then subtracts the number of substitutions that are actually observed. If there are fewer substitutions than expected (giving a higher score), then it indicates that the area is conserved (Davydov *et al.*, 2010). However it cannot calculate scores for INDELs. In this analysis a GERP++ score >2 was defined as conserved.

Amino acid sequence alignment was performed using UCSC Vertebrate Multiz Alignment & Conservation track (Blanchette *et al.*, 2004). Species, UCSC version given in brackets, used in the alignment were human (hg19), chimp (panTro2), orangutan (ponAbe2), mouse (mm9), rat (rn4), rabbit (oryCun2), dolphin (turTru1), dog (canFam2), opossum (monDom5), platypus (ornAna1) and zebrafish (danRer6). The symbol '-' is used when there is no amino acid in the aligned sequence. This could be due to an insertion or deletion in one of the species compared.

2.6.6.3. Variant Database

The Single Nucleotide Polymorphism Database (dbSNP) is a public collection of sequence variants (Sherry *et al.*, 2001, Smigielski *et al.*, 2000). The database accumulates variant data from large scale sequencing projects, such as 1000 genome project . The majority of the variants listed are missense mutations, compared to INDELs, occupying an estimated 99.77% of the database (Sherry *et al.*, 2001).

2.6.6.4. Population databases

Population databases have several uses in WES data analysis. Firstly it provides a full compendium of sequence variants in the region of interest. This could provide information as to whether there are other previously described mutations with clinical significance. Secondly it can be used to calculate the minor allele frequency (MAF). This is the percentage at which a variant occurs within a population and thus whether it was a rare or common change.

1000 genome (1000g) project is a database containing the sequence variants produced from a combination of exome and genome sequencing projects, from 14 populations (Abecasis *et al.*, 2012). Exome Variant Server (EVS) is another population database, containing whole exome data generated from the NHLBI GO Exome Sequencing Project (ESP6500). The dataset contains exomes from 2203 African American and 4300 European American unrelated individuals.

2.6.6.5. Splice site calculation

A web-based program was used to predict the splice site score of a given sequence . It is based on how similar the sequence matches the mammalian consensus sequences (Shapiro and Senapathy, 1987, Carmel *et al.*, 2004). The consensus sequence for the 3'

splice site is ‘ttttttttttcag/G and for the 5’ it is ‘CAGgtaagt’, where uppercase indicates exonic nucleotides and lower case indicates intronic nucleotides. The maximum score available for both sites is 100, indicating ‘best fit’ with the consensus sequence.

2.6.6.6. Phyre2

The crystal structure of some proteins investigated here (MCP, INF2, C9 and DGKε) were unknown, therefore Phyre2 was used. This is a server that predicts the approximate protein structure of an inputted amino acid sequence (Kelley and Sternberg, 2009), by comparing it to homologous proteins. This structure could then be used to model sequence variants and thus predict a potential effect on protein function. All proteins were run through the software in the ‘intensive’ modelling mode.

Protein	NCBI Reference sequence	Amino acids entered
MCP	NP 722548	35-285
INF2	NP 071934.3	1-250
C9	NP 001728.1	1-559
DGKε	NP 003638.1	1-567

Table 6 Protein sequences used in Phyre2.

2.6.6.7. PyMOL

Three-dimensional protein structures were manipulated using PyMOL (Schrödinger). Files were either generated using Phyre2 or downloaded from Protein Data Bank (Berman *et al.*, 2000). Protein domain boundaries for INF2 and DGKε were taken from Pfam (Finn *et al.*, 2014).

2.7. SDS PAGE

This was performed with the help of Geisilaine Soares dos Reis Araujo, Edwin Wong and Holly Anderson. Reducing sample buffer (Thermo Scientific, 39000) or non-reducing sample buffer (Thermo Scientific, 39001) was mixed with protein, diluted to appropriate concentration. To remove albumin and IgG from patient serum, which make up >70% of serum proteins, the Pierce antibody-based albumin/ IgG removal kit (Thermo Scientific, 89876) was used. This was necessary to try and visualise proteins of interest that ran at the same size, such as FHR4/5. Samples were heated at 95°C for 5 minutes and then centrifuged at a maximum speed of 13,200 rpm for 2 seconds. Tris-glycine gels (Novex, Life Sciences. 1.0mm x 10 well (10%- EC6075BOX, 6%- EC6065Box and 16%- EC6045Box)) and the XCell SureLock mini-cell (Novex, Life

Technologies. EI0002) were set up according to manufacturer's instructions (Life-Technologies, 2012) using 1x running buffer (25mM Tris base, 192mM Glycine, 0.1% SDS pH 8.3) (Life Technologies, LC2675-5). 15µL of sample was loaded per well and 7µL of molecular weight ladder (Thermo Scientific, 26619) or (Biolabs, P7708s). This was connected to a PowerPac (Bio-Rad. 300V, 400mA, 75W) and run for 90 minutes at 125 volts.

2.8. Western blotting

2.8.1. Protein transfer

XCell SureLock Mini-Cell was set up for gel transfer, as written in the manufacturer's instructions (Life-Technologies, 2009) for Novex Tris-glycine gels onto a nitrocellulose membrane (Invitrogen, Life technologies. LC2001). Gels were transferred for 90 minutes at 25 volts using 1x Tris-glycine transfer buffer (12mM Tris base, 96mM Glycine, pH 8.3, 20% Methanol), which was cooled to approximately 4°C before use. After transfer the nitrocellulose membrane and Tris-glycine gels were washed with H_2O. To test if protein had transferred, the membrane was stained using Ponceau S solution (Sigma, P7170) and the gel with Coomassie G250 Stain (Bio-Rad, 161-0786). The membrane was then blocked with blocking solution, 5% non-fat milk powder in 1x TBST (50mM Tris.HCl, pH 7.4, 150mM NaCl, 0.05% Tween 20), for 1 hour at room temperature or 4°C overnight. A rotating table was used to ensure even coverage of solution over the membrane. Fresh blocking solution was then added, containing the primary antibody at the desired concentration. This was left for 1 hour at room temperature. The liquid was discarded and the membrane was washed for one 15 minute wash and two 5 minute washes, using fresh 1x TBST. Fresh blocking solution containing appropriate concentrations of secondary antibody was then added and left for 1 hour at room temperature. The liquid was discarded and the wash steps repeated, as stated previously. Table 7 shows the conditions used for western blotting.

Protein target	FH	FHRs 1,2 and FHL	FHRs 4/5	FH	FH	FH
Size range	150kDa	25-47kDa	50-70kDa	150kDa	150kDa	150kDa
Reduced/ Non-reduced	Reduced	Reduced	Reduced	Non- reduced	Non- reduced	Non- reduced
Serum preparation	ND	ND	IgG and Albumin removed	ND	ND	ND
Serum dilution	1:1500	1:200	1:60	1:100	1:100	1:100
Gel Percentage	6%	16%	10%	6%	6%	6%
Primary antibody	Polyclonal goat anti-FH (Calbiochem, 341276)	Polyclonal goat anti-FH (Calbiochem, 341276)	Polyclonal goat anti-FH (Calbiochem, 341276)	Mouse anti-CCP20 of FH (L20/3) (Santa Cruz Biotechnology. SC-47686)	Mouse anti- SCR5 of FH (OX24) (Hybridoma cell line-Sigma-Aldrich, 00010403)	MBI6 and MBI7. Mouse anti-402 H or Y in CCP7 of FH. (Donated by Claire Harris, Cardiff University)
Concentration	1:4000	1:4000	1:4000	1:2000	1:2000	1:2000
Secondary antibody	Rabbit anti-goat HRP (Calbiochem, 401515)	Rabbit anti-goat HRP (Calbiochem, 401515)	Rabbit anti-goat HRP (Calbiochem, 401515)	Sheep anti-mouse IgG HRP (Jackson Immuno Research, 515-035-071)	Sheep anti-mouse IgG HRP (Jackson Immuno Research, 515-035-071)	Sheep anti-mouse IgG HRP (Jackson Immuno Research, 515-035-071)
Concentration	1:8000	1:4000	1:4000	1:4000	1:4000	1:4000

Table 7 Western blotting conditions.

ND= Not done. HRP=Horseradish peroxidise.

2.8.2. Protein visualisation

SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, 34077) was added to the membrane and left for 1 minute. The membrane was then dried, covered in cling film and place inside of the hypercassette (Amersham Biosciences, RPN 13642). Film (Sigma Aldrich, 1611342) was then placed over the membrane and the hypercassette was closed, leaving sufficient exposure time to yield bands. Finally the film was developed (Konica Minolta Medical + Graphic, Inc. SRX-101A).

2.8.3. Membrane restoration

If another antibody was to be tested on the same membrane, Restore Western Blot Stripping Buffer (Thermo Scientific, 21059) was used. Membrane was first washed in H_2O and then incubated in stripping buffer for 10 minutes at room temperature, then washed in 1x TBST for 5 minutes. To ensure complete antibody removal, the membrane was developed as described previously.

2.9. Immunoaffinity Chromatography

Affinity chromatography is a technique that was used to separate molecules according to a specific factor, such as pH, size, charge or binding specificity. Immunoaffinity chromatography utilises antibody-specificity to separate proteins of interest. This method was used to separate the FH and FH/FHR3 hybrid protein species from patient serum. This technique was carried out on an Äktapurifier UPC 10 (GE Healthcare, 28-4062-68).

To separate FH and FH/FHR3 from serum a column was synthesised containing OX24, which is a monoclonal antibody specific to CCP 5 of FH protein and FHL, shown in Section 3.2.3.1, Figure 3-10. The conjugation of OX24 to the column would allow the proteins of interest to bind, separating them from other serum proteins. These proteins could then be eluted for further analysis.

2.9.1. Tissue culture of OX24 hybridomas

OX24 antibody was produced by a B cell hybridoma cell line (Sigma-Aldrich, 00010403). Hybridomas were cultured at 37°C and 5% CO_2 using Roswell Park Memorial Institute (RPMI) medium supplemented with 10% low immunoglobulin foetal calf serum, 2mM L-glutamine, 0.1units/mL penicillin, 0.1mg/mL streptomycin,

1mM non-essential amino acids, 1mM sodium pyruvate (all PAA) and 50uM β-mercaptethanol (Sigma-Aldrich). To obtain the OX24-containing cell media, the cell suspension was centrifuged at 300xg for 2 minutes. The supernatant was removed and stored at -20°C.

2.9.2. Cryopreservation of cells

To prepare cells for cryopreservation the cell suspension was centrifuged at 300xg for 2 minutes. Once the supernatant was removed the cell pellet was resuspended in Recovery™ Cell Culture Freezing Medium (Life Technologies, 12648-010) and aliquoted into Cryogenic Tubes (Thermo Scientific, 377267). The tubes were then placed into a Mr. Frosty™ Freezing Container Freezing Container (Thermo Scientific, 5100-0001) with 2-Propanol (VWR, 20842.323) and stored at -80°C. After approximately 24 hours the cryovials were transferred to liquid nitrogen storage.

2.9.3. OX24 purification

Supernatant from the OX24 hybridomas, containing the OX24 antibody, was mixed with PBS (0.01M phosphate buffer, 0.0027M potassium chloride and 0.137M sodium chloride, pH 7.4) and loaded onto a 5 mL HiTrap Protein G HP column (GE Healthcare, 17-0405-01). The protein G within the column has a high affinity for the Fc region of the IgG (GE-Healthcare, 2015), allowing OX24 separation from the cell media. OX24 was then eluted with 0.1M Glycine-HCl pH2.7 into 0.5mL fractions containing 80µL 1M Tris base pH9.0 to neutralise the eluted antibody. Fractions containing the OX24 antibody were buffer exchanged using Vivaspin 2 PSE, 5000MWCO (Fisher Scientific, VS0211) into 1mL coupling buffer (0.2M $NaHCO_3$, 0.5M NaCl pH8.3).

2.9.4. Construction of an OX24 column

The OX24 column was made using a 1mL HiTrap NHS-activated HP (GE Healthcare, 17-0716-01). The ester- activated N-hydroxysuccinimide (NHS) ligands on the column agarose matrix react with amino groups found on the antibody. This reaction produces amide bonds that stably bind the antibody to the column matrix, shown in Figure 2-6.

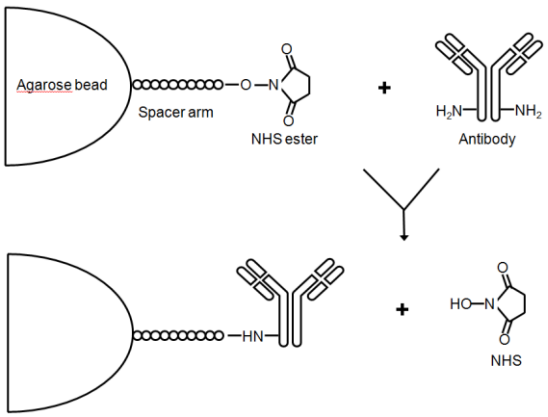

Figure 2-6 Diagram showing immobilisation of antibody to column.

Column matrix is formed of agarose beads coated in ester-activated NHS ligands, attached via a 10 atom spacer arm. The ester groups then reacts with a free amino group on the antibody, forming an amide bond and releasing NHS (Thermo-Scientific).

Formation of the antibody-conjugated column was performed manually, according to the manufacturer's instructions for the HiTrap NHS-activated HP column (GE-Healthcare, 2013b). During ligand coupling the antibody was flowed across the column, between two syringes at either end of the column. This was carried out at room temperature for 1 hour. After coupling the column was washed with 3 column volumes of coupling buffer and the eluate was collected, which was used to calculate the coupling efficiency.

2.9.5. Calculating coupling efficiency

To calculate the coupling efficiency, two methods were used to determine protein concentration of pre- and post- coupling solutions. First using a NanoDrop and second using a bicinchoninic acid (BCA) protein assay (Thermo Scientific, 23225). The PD-10 desalting columns (GE Healthcare, 17-0851-01) were used to remove free amines release during the coupling reaction, which may affect the BCA assay and NanoDrop reading. The PD-10 columns were used according to the manufacturer's gravity protocol (GE-Healthcare, 2013a). Absorbance of protein on the NanoDrop was measured at 280nm. Coupling efficiency was then calculated, as written in the manufacturer's instructions (GE-Healthcare, 2013b), to be approximately 90%. The BCA protein assay was carried out according to manufacturer's microplate protocol (Thermo-Scientific, 2013). The pre-coupling (diluted 1:10) and post coupling (after PD-10 purification) antibody samples were tested. Absorbance was measured using a Thermo Labsystems Multiskan Ascent (Thermo Scientific), with a 570nm filter and

Ascent Software version 2.6 (Thermo Scientific, 5185430CD). Table 8 shows the standard curve for the blank-corrected absorbance for BSA standards.

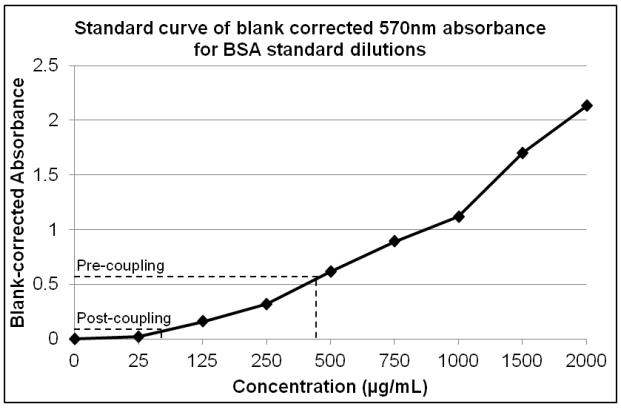

Table 8 Standard curve BSA standard dilutions.

Graph showing the standard curve for blank corrected 570nm absorbance, for BSA standard dilutions.

The average, blank-corrected absorbance for the pre-coupling sample was 0.56 and post coupling sample was 0.04. If these values are drawn on the curve, it gives an approximate sample concentration for the diluted pre-coupling sample of 450µg/mL (pure sample estimated to be 4500µg/mL) and post-coupling sample has a concentration of approximately 40µg/mL. This gives an approximate coupling efficiency of 99.1%.

2.9.6. Serum purification using OX24 column

Serum was diluted with an equal volume of binding buffer (PBS) and filtered using an Acrodisc Syringe filter (Pall Corporation, PN4612). The OX24 column was loaded onto the Äktapurifier and the diluted serum was loaded by hand or using a 10mL Superloop (GE Healthcare, 18-1113-81). Proteins of interest were eluted using 0.1M Glycine pH2.7 into 0.5mL fractions containing 50µL 1M Tris pH8.0, to neutralise the pH of eluted protein. Fractions containing protein of interest were then pooled and concentrated to an appropriate volume. This step was carried out with the assistance of Holly Anderson and Edwin Wong.

2.10. Mass spectrometry

Approximately 500µL of fresh serum was purified as described in section 2.9.6. The eluted material was pooled and concentrated to 30µL. This was then run on a 6% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) and stained

with Coomassie G250. Mass spectrometry was performed by Achim Truemann, where protein bands were excised from the gel and reduced with 10mM DTT (Sigma-Aldrich) in 100mM NH_4HCO_3, alkylated with 50mM iodoacetamide (Sigma-Aldrich) in 100mM NH_4HCO_3. The proteins were then digested in gel, with 230ng modified trypsin (Promega) in 50mM NH_4HCO_3, 1mM $CaCl_2$. Peptides were extracted from the gel pieces and the digests were analysed by LCMSMS using a Dionex U3000 nano-HPLC system (Thermo Scientific), coupled to an LTQ Orbitrap LTQ XL™ ETD (Thermo Scientific) mass spectrometer. Peptides were separated on a 25 cm x 75 µm PepMap column (Thermo Scientific) using a 37 min water acetonitrile gradient (0.05% Formic acid). Precursor ions were detected in positive mode at 350-1600 m/z with a resolution of 60,000 (at 400 m/z) and a fill target of 500,000 ions and a lockmass was set to 445.120023 m/z. The five most intense ions of each MS scan (with a target value of 10,000 ions) were isolated, fragmented and measured in the linear ion trap. Peaklists in the Mascot generic format (*.mgf) were generated using msconvert ((Kessner *et al.*, 2008, Chambers *et al.*, 2012)) and the ensembl human genome (GrCh37.66) was searched using X!Tandem and the gom interface (version 2013.09.01.1(Craig and Beavis, 2004)), with carbamidomethyl set as a fixed modification and Met oxidation set as a variable modification. Two refinement steps were included in the search to include deamidation and methylation artefacts as well as protein phosphorylation, acetylation, dehydration of threonine and serine and carbamidomethylation of glutamine, histidine, aspartic acid, glutamic acid and lysine. The protein level false positive rate (as defined in: http://wiki.thegpm.org/wiki/False_po sitive_rate) in each band was below 1%.

2.11. Cell Surface Decay Haemolytic assay

Haemolytic assays are *in vitro* models that measure complement activity at the cell surface (Harris, 2000). The decay acceleration assay examines the ability of a FH to accelerate the decay of the AP convertase on a sheep erythrocyte. This requires FH to bind to GAGs on the surface of the erythrocyte, so that it is in close proximity to the convertase. The decay of the convertase on the erythrocyte surface will prevent the cleavage of C3, which is added back to the system. In doing so there will be less C3b formation and thus reduced terminal complement pathway activity. Fewer MACs will be produced and as a result there will be less lysis of the erythrocytes. If convertase decay fails to take place, due to an abnormality of FH, then the convertase will remain

intact. When serum is added to the system, C3 will be cleaved into C3b and activate terminal complement activity, which will lead to MAC formation and cell lysis.

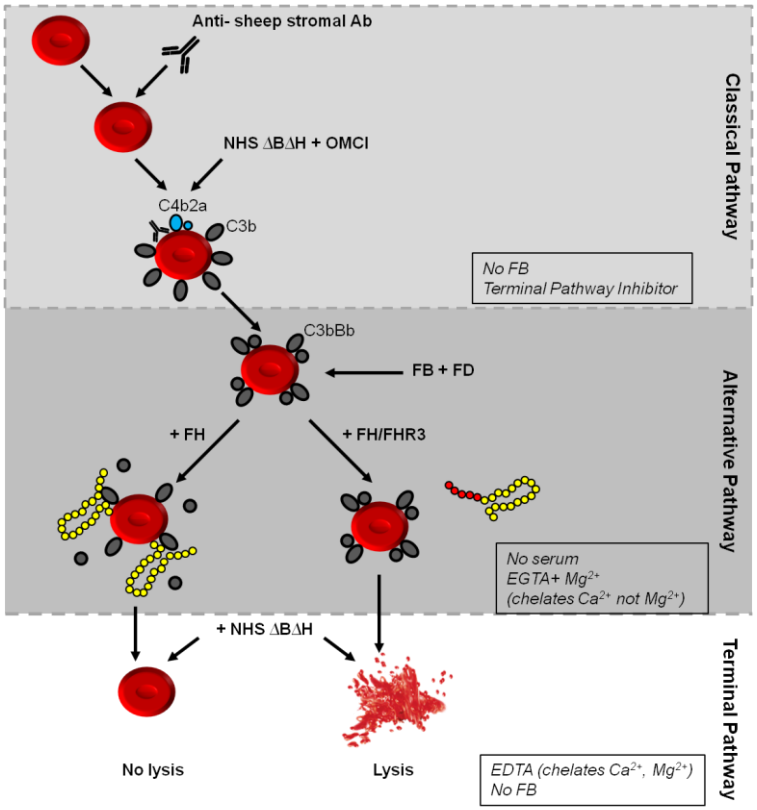

Figure 2-7 Schematic diagram of the cell surface decay haemolytic assay.

Sheep erythrocytes were sensitized by mixing with anti-sheep stromal antibodies. This led to CP activation and CP C3 convertase (shown in blue) formation. This cleaved C3 into C3b (shown in grey) upon the addition of NHSΔBΔH. The absence of FB stopped AP activation, absence of FH prevented decay acceleration and the presence of a terminal pathway inhibitor prevented the cells from being lysed. FB and FD were then added, which led to the formation of the AP convertase. This is done in the presence of EGTA, which preferentially chelated Ca^{2+} ions, inhibiting CP activity. The FH species were then added (shown in yellow), wild type FH would accelerate the decay of the AP convertase and so will protect cells from lysis upon the addition of NHSΔBΔH. The FH/FHR3 would not accelerate the decay of the AP convertase and therefore the addition of NHSΔBΔH would activate the terminal pathway leading to MAC formation and thus cell lysis. This was carried out in EDTA, using NHSΔBΔH to prevent CP and AP activity, which may skew the results.

2.11.1. Serum preparation

Fresh patient and control FH species were purified using immunoaffinity chromatography with an OX24 column, as described in section 2.9.6. The eluted material was pooled and concentrated using Vivaspin 2 PSE, 5000 MWCO columns (Fisher Scientific, VS0211). Normal human serum depleted of FB and FH (NHSΔBΔH), was prepared by immunoaffinity chromatography, where serum was run across JC1

(FB-specific antibody) and OX24 antibody-conjugated columns in series, using the method described in section 2.9.6. NHSΔBΔH and purified FB and FH were collected and pooled, for use in the cell surface decay haemolytic assay.

2.11.2. Protein quantification

FH species were purified from approximately 500µL of fresh patient serum and 2mL fresh control serum, as written in section 2.9.6. The OD at 280nm for control FH was 5.975mg/mL and patient FH mixture (FH and FH/FHR3) was 0.1535mg/mL. Molarities were calculated using Beer-Lambert Law: A= εLc, (where A= Absorbance, ε= Molar extinction coefficient, L= light path and c= Concentration). The extinction coefficients were determined using ProtParam, assuming all pairs of cysteines form cystines (Gasteiger *et al.*, 2005). The protein sequences for FH (NP_000177) and FHR3 (NP_066303) were obtained from NCBI Protein (NCBI Resource Coordinators, 2014). The leader sequences for FH (MRLLAKIICLMLWAICVA) and FHR3 (MLLLINVILTLWVSCANG) were excluded from the analysis. The extinction coefficient for FH was 246800M.cm^{-1}. To calculate the extinction coefficient for FH/FHR3, the first 1026 amino acids from FH and the 312 amino acids from FHR3 were used. This produced an ε of 272170M.cm^{-1}. This would ensure conservative estimates of the molar ratios and as such, differences demonstrated in the functional assays were an underestimate of the loss of function.

2.11.3. Assay methodology

The cell surface decay haemolytic assay was performed as previously described (Wong *et al.*, 2014, Tortajada *et al.*, 2009), using a method that had been optimized by Edwin Wong. I performed this experiment, with the assistance of Edwin Wong. Sheep erythrocytes were sensitised (EA) by coating cells with antibodies. To do this 2% washed sheep erythrocytes (TCS biosciences, SB068) were mixed with warmed Complement Fixation Diluent (CFD) (Oxoid, BR0016), containing 1:4000 anti-sheep erythrocyte stromal antibodies (Sigma, S1389). The mixture was then incubated at 37°C for 30 minutes and washed using CFD and resuspended to 2% (v/v).

C3b was then deposited onto EA cells (EAC3b). To do this 8% NHSΔBΔH was mixed with 1:1666 *Ornithodoros moubata* complement inhibitor (OmCI) (gift from Varleigh Immuno Pharmaceuticals) in 2% (v/v) EA. This was incubated for several minutes at

room temperature, before being added to an equal volume of CFD. The mixture was then incubated at 37°C for 20 minutes. EAC3b was washed with AP Buffer (5mM sodium barbitone, pH7.4, 150mM NaCl, 7mM $MgCl_2$, 10mM EGTA) and resuspended to give 2% (v/v) EAC3b cells. A convertase was formed on the sheep erythrocytes by adding 2% EAC3b cells to an equal volume of AP Buffer that contained 0.04µM FB (1:400) (prepared in house) and 0.2µg/mL FD (1:500) (Complement Technology, A136). This was incubated for 15 minutes at 37°C.

The assay was performed in round bottom microtiter plates (USA Scientific, 5665-0161). A 1:2 titration series (0.04µM – 80nM) of OX24-purified FH from control or patient was made in 50µL PBS/ 20mM EDTA. Three no FH controls (0.1% (v/v) Triton X-100 (Sigma-Aldrich, T8787), with serum or without serum) were done in parallel. 100µL of convertase mixture was added to each well and the plate was incubated at room temperature for 15 minutes, to allow decay acceleration to take place. 50µL 8% (v/v) NHSΔBΔH in PBS/20mM EDTA was added to wells (except no FH and no serum, or no FH and Triton controls), to initiate lysis. The plate was covered with adhesive film (VWR, 391-1250) and incubated at 37°C for 30 minutes, then centrifuged at 1200rpm for 3 minutes. 100µL supernatant was then transferred into a flat bottom microtiter plates (Ratiolab, 6018211). The amount of haemoglobin released was measured on a Thermo Labsystems Multiskan Ascent (Thermo Scientific) using a 420nm filter and Ascent Software version 2.6 (Thermo Scientific, 5185430CD). All values were blank-corrected, using the mean A_{420} of no serum/no FH controls. Maximal lysis was achieved by adding NHSΔBΔH to no FH wells (buffer only). Percentage of inhibition of lysis in the presence of increasing concentrations of FH was defined as: (A_{420}[buffer only]-A_{420}[FH]) / A_{420}[buffer only] *100%.

2.11.4. Sigma Plot

Results of the cell surface decay haemolytic assay were plotted using Sigma Plot (Systat Software Inc).

Chapter 3: Genetic abnormalities in *CFH* and *CFHRs*

3.1. Introduction

Previous studies have shown that the most common genetic abnormality in aHUS is in *CFH,* occurring in approximately 30% of the affected population (Caprioli *et al.*, 2006, Loirat and Fremeaux-Bacchi, 2011, Maga *et al.*, 2010). *CFH* is found in the RCA cluster, in close proximity to the genes encoding the 5 Complement Factor H related proteins (*CFHRs*) (Jozsi and Zipfel, 2008). This area contains several LCRs that can cause genomic instability and thus predispose this region to gene rearrangements and conversions (Skerka *et al.*, 2013).

3.1.1. Gene conversions

Gene conversions are formed when genetic sequences are transferred between two sites, which share a high degree of sequence homology (Chen *et al.*, 2007). Domains 18-20 of FH and domains 3-5 of FHR1 have been demonstrated to be highly homologous (Jozsi and Zipfel, 2008). Indeed CCPs 18 of FH and 3 of FHR1 differ by 3 amino acids, domains 19 and 4 share identical amino acid sequences, although there is a difference at 1 nucleotide and finally CCPs 20 and 5 differ by 2 amino acids.

Heinen *et al.* (2006) identified aHUS patients with sequence variants in CCP20 of FH; c.3572C>T, p.S1191L, c.3590T>C, p.V1197A or both. These variants correspond respectively to the amino acids L290 and A296, found in CCP 5 of FHR1. They demonstrated that these variants arose due to gene conversion between exon 6 of *CFHR1* and exon 23 of *CFH*. Functional analysis demonstrated that recombinant proteins containing S1191L and V1197A (shown in Figure 3-14A), had a reduced affinity for C3b and C3d (Heinen *et al.*, 2006) and Ferreira *et al.* (2009) demonstrated that this led to impaired cell surface complement regulation.

3.1.2. Gene rearrangements

Venables *et al.* (2006) identified aHUS patients possessing a *CFH/CFHR1* hybrid gene, the result of NAHR. This produced a protein containing CCPs 1-18 from FH and CCPs 4-5 from FHR1, shown in Figure 3-14D. This protein was indistinguishable in structure to the protein described by Heinen *et al.* (2006). This genomic rearrangement was not detected by Sanger sequencing, instead MLPA revealed a deletion of exons 21-23 of

CFH. The patient's cDNA was analysed and demonstrated to contain an abnormal transcript. The patient had normal serum FH levels, indicating that the abnormal protein was secreted normally.

Maga *et al.* (2011) then described an alternative *CFH/CFHR1*, also the result of NAHR. MLPA analysis of the *CFH/CFHR* gene cluster demonstrated a deletion ranging from exon 23 of *CFH* up to and including exon 5 of *CFHR1*. The resultant protein was identical in size to the FH/FHR1 reported by Venables *et al.* (2006), however it consisted of CCPs 1-19 of FH and CCP 5 of FHR1 (shown in Figure 3-14C). The protein was secreted, although functional analysis indicated that the protein lacked cell surface binding activity.

Two reverse hybrids have been reported where the C-terminal domains of FHR1 are replaced by the C-terminus of FH. The first was reported by Eyler *et al.* (2013) where MLPA identified a duplication of exons 21 of *CFH* to exon 4 of *CFHR1* and the break point was located in *CFHR1* intron 4. NAHR produced a *CFHR1/CFH* hybrid gene encoding the first 3 CCPs of FHR1, followed by CCPs 19 and 20 of FH (shown in Figure 3-14F). The second reverse hybrid was reported by Valoti *et al.* (2015). This FHR1/FH hybrid was identical in size to the Eyler hybrid, although it contained CCPs1-4 of FHR1 and CCP20 of FH, shown in Figure 3-14E. Functional analysis demonstrated decreased complement regulation at the surface of cells. The addition of C-terminal CCPs from FH was predicted to allow the hybrids to bind to the cell surface. However their lack of N-terminal regulatory domains (CCPs 1-4 in FH) meant that they were unable to regulate complement. Therefore they were hypothesised to competitively bind to the cell membrane and prevent FH-mediated complement regulation (Valoti *et al.*, 2015).

Francis *et al.* (2012) discovered a *CFH/CFHR3* hybrid that occurred as a result of a deletion of the end of *CFH* (exon 23), causing the gene to splice into the downstream gene, *CFHR3*. This was the first gene rearrangement to have occurred by an alternative repair mechanism, MMEJ. The resultant protein contained CCPs 1-19 from FH and all 5 CCPs of FHR3, shown in Figure 3-14G. The protein was shown to be secreted normally; however it was unable to regulate complement at the cell surface due to the loss of CCP20 of FH, which prevented FH from binding to GAGs found on the cell surface. This was hypothesised to lead to onset of disease.

3.1.3. Chapter aims

In this chapter the analysis of copy number variation in the RCA cluster is described, which was performed in all unsolved familial aHUS cases and selected others in the Newcastle cohort prior to further genetic analysis. All 28 families analysed in this project were analysed for CNVs using MLPA and western blotting, if there was DNA and sera available. The results can be found in the Supplementary data.

A novel *CFH/CFHR3* hybrid gene is described, with evidence of a secreted hybrid protein and functional analysis demonstrating the consequences of the altered protein structure. Gene rearrangements and conversion events in *CFH* and *CFHRs* are then examined and put into the context of all rare sequence variants reported in *CFH* in the Newcastle cohort.

3.2. Hybrid *CFH/CFHR3* gene

3.2.1. Clinical history

The patient presented at 8 months of age with no previous family history of disease, see Figure 3-1 for pedigree. The patient had a diarrhoeal illness and on further investigation, was found to have a creatinine level of 52μmol/L, MAHA was seen on a blood film and the LDH level was 5747U/L. His stool culture was positive for *E. coli* O157:H7, leading to a diagnosis of STEC HUS. Renal replacement therapy was not required and the patient did not receive plasma exchange. He was discharged 2 weeks later with a creatinine of 107μmol/L.

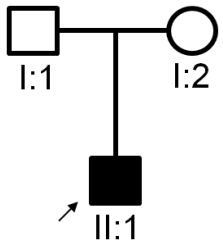

Figure 3-1 Family pedigree.

The affected proband is indicated by an arrow.

Two weeks later the patient was readmitted with an upper respiratory tract viral infection. Blood tests revealed a creatinine of 141μmol/L, LDH 2398U/L and a blood film demonstrated MAHA. Serum levels of C3 (1.05g/L), C4 (0.30g/L), FH (0.51g/L)

and FI (71mg/L) were found to be within the normal range and ADAMTS13 activity was normal. A renal biopsy revealed thrombi in microcapillaries of the glomerulus confirming a diagnosis of aHUS, shown in Figure 3-2.

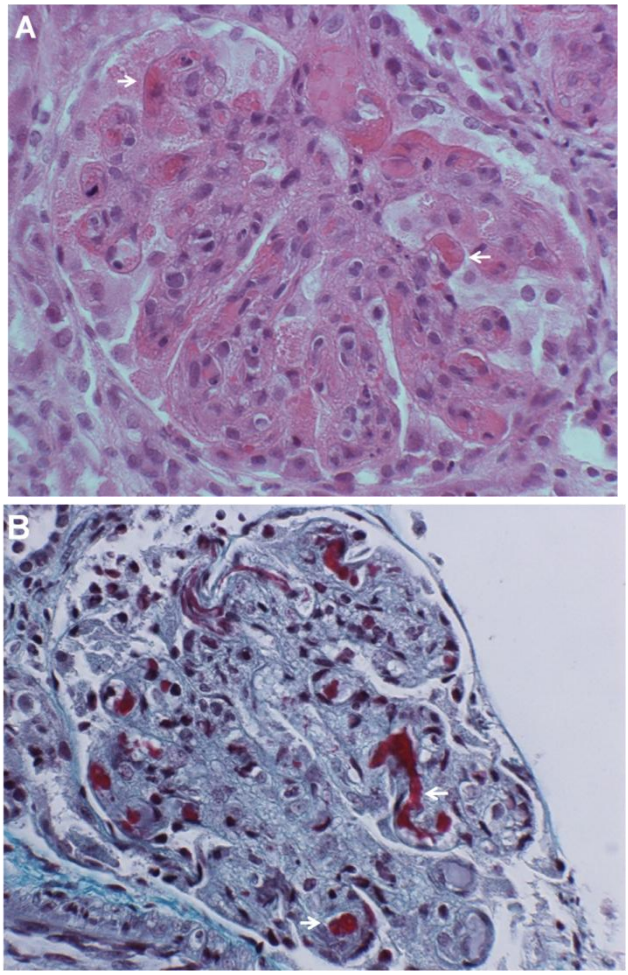

Figure 3-2 Histological sections of patient renal biopsy.

A- Haematoxylin and eosin staining (x400 magnification). B- Masson trichrome staining (x400 magnification). Arrows indicate thrombi in glomerulus vasculature. Reprinted with permission from Challis et al. (2015).

Plasma exchange was initiated and the patient required three haemofiltration sessions. The creatinine levels returned to a baseline of ~100μmol/L and a weekly regime of plasma exchange was started. This continued for three and a half years before switching to eculizumab. For three years the patient has not had any disease recurrence, with a current creatinine level of 200μmol/L.

3.2.2. Genetic analysis

3.2.2.1. Sanger sequencing

Diagnostic testing of aHUS-associated genes (*CFH*, *CFI*, *CFB*, *MCP* and *C3*) by the Northern Genetics Service, revealed no mutations. However it was noted that the patient was heterozygous for the polymorphism in *CFH*, c.1204C>T, p.Y402H (rs1061170).

3.2.2.2. MLPA

To detect potential CNVs in *CFH* and the *CFHRs*, which would not be detected by Sanger sequencing, MLPA screening was carried out. Initial diagnostic screening was performed by the Northern Genetic Service, of *CFH*, *CFHR1*, *CFHR2*, *CFHR3* and *CFHR5*. The MLPA results, shown in Figure 3-3, detected a normal number of gene copies for all exons of all 5 genes, except for exon 23 of *CFH*, where the copy number was below the threshold (0.8), suggesting a deletion. The proband was also tested using the MLPA probes designed in house for *CFH* and *CFHR5*, by Geisilaine Soares dos Reis Araujo, which supported these findings (shown in the Supplementary data).

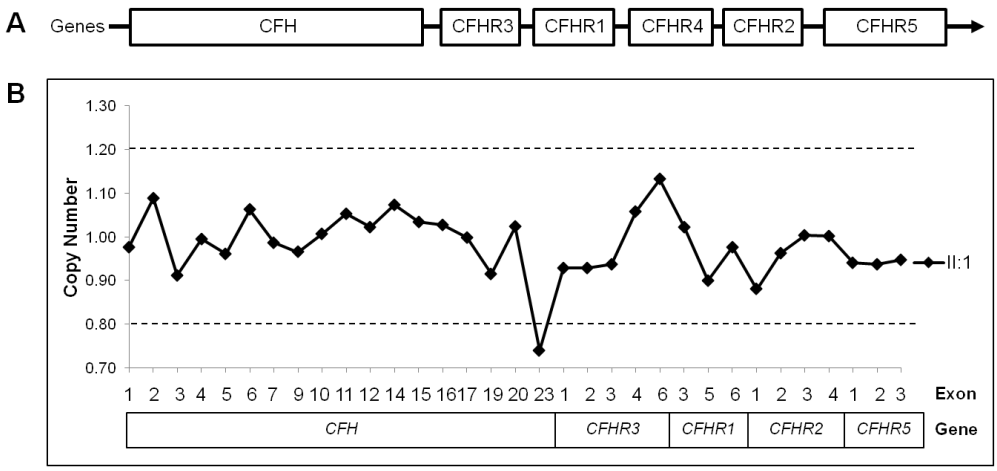

Figure 3-3 MLPA screening results

A- Gene positions of CFH and CFHRs on Chromosome 1. B- MLPA trace of CFH, CFHRs 1,2,3 and 5 for proband (II:1). Reprinted with permission from Challis et al. (2015).

3.2.2.3. Break point analysis

To confirm the location of the deletion, Sanger sequencing was carried out by the Northern Genetic Service. Patient genomic DNA was amplified using a combination of primers positioned throughout the last exons of *CFH* and into the 3'untranslated region (UTR). Only when using a forward primer in exon 20 of *CFH* and a reverse primer in

CFH 3' UTR, was a product made. Sequencing of this PCR product revealed that the break point, shown in Figure 3-4, was located 3' of exon 20. This led to the deletion of exons 21 to 23 of *CFH*, comprising of a 6.3kb loss of sequence. The parents of the proband were then tested, but the deletion was not found. Maternal and paternal status was confirmed by the Northern Genetics Service, using microsatellite testing. This led to the conclusion that this was a *de novo* event.

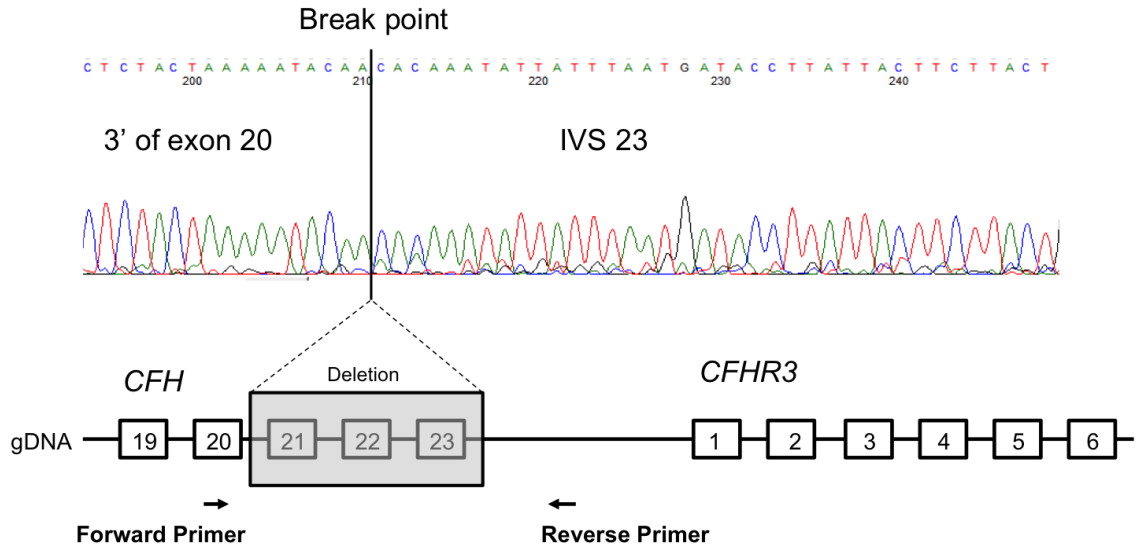

Figure 3-4 Sequencing chromatogram of break point.

Top, sequencing chromatogram. The breakpoint in CFH is demonstrated with a line. Below, deletion is shown by a grey box. Primer positions are marked by arrows. Reprinted with permission from Challis et al. (2015).

It was hypothesised that sequence homology between *CFH* and *CFHR3*, close to the position of the break point, may have predisposed this region to a genetic rearrangement. Therefore to test this, the genomic DNA (gDNA) sequence for the patient breakpoint was compared to the sequences for wild type *CFH* and *CFHR3*. A seven-nucleotide sequence 'AATACAA' was found to be identical in *CFH* and *CFHR3* adjacent to the break point, highlighted by a box in Figure 3-5.

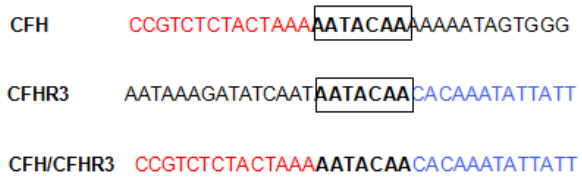

Figure 3-5 Breakpoint sequence homology.

Sequence overlay for CFH and CFHR3 and CFH/CFHR3 breakpoint. Red- sequence for CFH seen in CFH/CFHR3, Blue- sequence for CFHR3 seen in CFH/CFHR3. Microhomology sequence is boxed and in bold text. Reprinted with permission from Challis et al. (2015).

It was predicted that the rearrangement occurred by intrachromatidal recombination, due to MMEJ (Ma *et al.*, 2003). MMEJ is a DNA repair mechanism that occurs at DSBs and uses LCRs in close proximity, to align the broken fragments (Chen *et al.*, 2010). It was hypothesised that this mechanism, incorrectly aligned the 7bp sequence from the two genes, within a chromosome. As a result it caused exons 21- 23 of *CFH* to be excised and deleted, demonstrated in Figure 3-6.

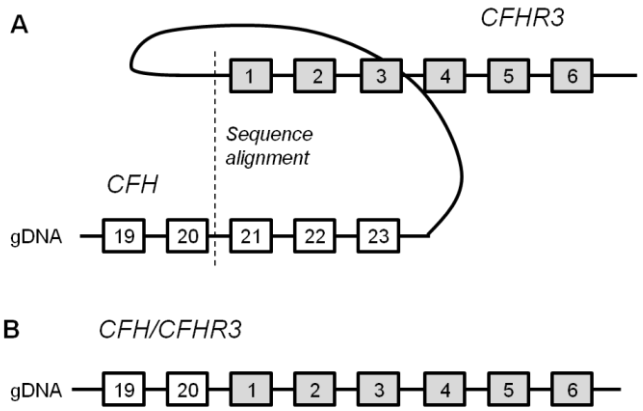

Figure 3-6 Diagram showing intrachromatidal MMEJ.

A- Approximate position of aligned microhomologous sequences in CFH and CFHR3. B- The resulting hybrid CFH/CFHR3 gene containing exons 1-20 of CFH and exons 1-6 of CFHR3.

This deletion was shown not to affect the reading frame, but the loss of the terminal exons and 3' UTR of *CFH* was proposed to cause aberrant splicing of *CFH*. It was hypothesised that the intact splice donor site of *CFH* exon 20, would splice into the next available splice acceptor site. *In silico* software suggested that the splice acceptor site of exon 2 in *CFHR3* would be used, shown in Figure 3-7. This was predicted to result in the formation of an inframe *CFH/CFHR3* hybrid gene.

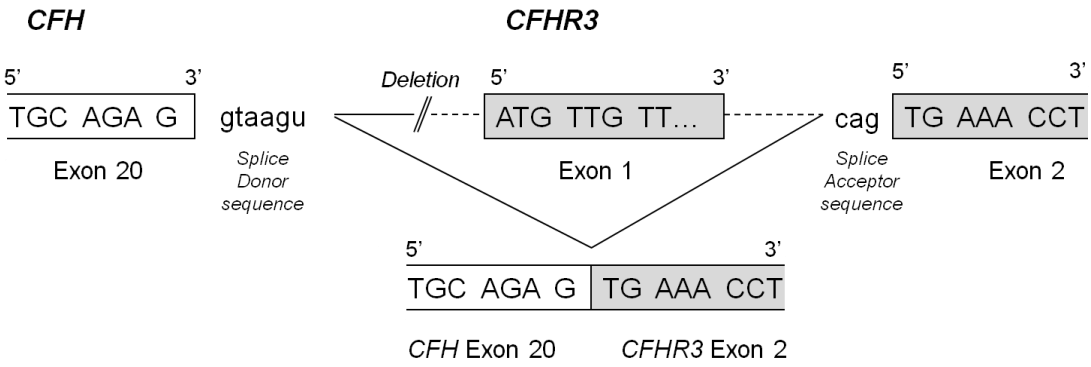

Figure 3-7 Schematic diagram of CFH/CFHR3 splicing.

The sequences for exon 20 CFH are in white and exons 1 and 2 of CFHR3 are in grey. Top- positions of exons on DNA. Bottom- predicted splicing of exon 20 CFH and exon 2 of CFHR3. Sequence reading frame remains unaltered in hybrid gene.

The mRNA product transcribed from the hybrid gene, was be predicted to contain all coding exons of *CFH*, except for 21-23, with the addition of exons 2-6 from *CFHR3*. To investigate if an mRNA transcript is produced from the hybrid gene and that it has the predicted structure, patient and control cDNA were analysed. The cDNA was amplified using primers, positioned across the junction between *CFH* and *CFHR3*. The forward primer was positioned in exon 20 of *CFH* and a reverse primer in exon 2 of *CFHR3*, shown in Figure 3-8. This design would only allow a product, with a predicted size of 108bp, to be made if there was a hybrid *CFH/CFHR3* present. PCR would not take place using control cDNA or patient gDNA. This was because control cDNA does not contain the binding sites for both primers as they would be present on different mRNA transcripts. Then if patient gDNA was present as a contaminant, the intronic region between *CFH* and *CFHR3* would be too large (approximately 27kb) to allow PCR to take place.

When patient and control cDNA was amplified in this experiment, a product was only seen in the patient, shown in Figure 3-8. This supported the hypothesis that the patient had a hybrid mRNA transcript, produced from the hybrid *CFH/CFHR3* gene. Insufficient quantities of cDNA meant that the PCR product was unable to be sequenced.

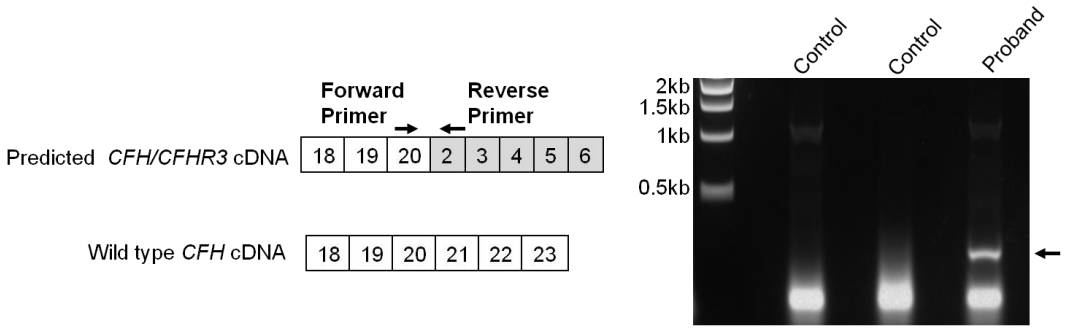

Figure 3-8 Confirmation of CFH/CFHR3 cDNA product.

Left- schematic diagrams showing the C-terminus of the predicted CFH/CFHR3 hybrid and wild type CFH cDNA transcripts. Positions of forward and reverse primers are indicated by arrows. Right- Image of agarose gel. PCR product seen in proband is indicated by an arrow. Reprinted with permission from Challis et al. (2015).

3.2.3. Proteomic analysis

3.2.3.1. Western blotting

From the genetic data obtained it was inferred that the hybrid *CFH/CFHR3* gene contained exons 1-20 from *CFH*, followed by the 5 exons of *CFHR3*. Exon 1 of *CFH* encodes the 5'UTR and leader peptide. CCP2 is encoded by 2 exons, exons 2-3. The remaining 18 CCPs of FH are encoded by one exon each. Exon 1 of *CFHR3* encodes the leader peptide and exons 2-6 encode the 5 CCPs. The resulting hybrid FH/FHR3 protein was proposed to contain the 17 N-terminal CCP domains from FH and all 5 CCPs of FHR3, shown in Figure 3-9.

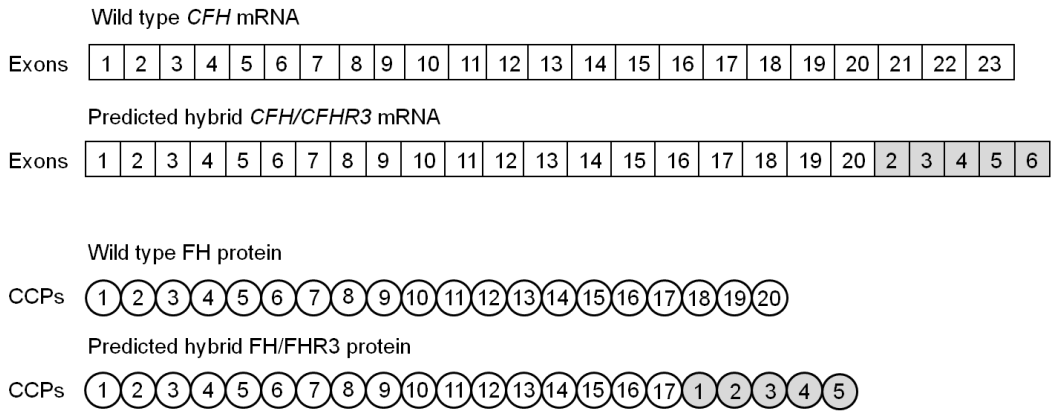

Figure 3-9 Predicted CFH/CFHR3 mRNA transcript and protein.

Top- mRNA transcripts of CFH and predicted CFH/CFHR3. Bottom- protein structure of FH and predicted FH/FHR3. Exons and CCPs derived from CFHR3 or FHR3 are grey. Exons and CCPs derived from CFH or FH are white.

In order to determine if the hybrid mRNA transcript led to the formation of a protein, patient serum was analysed by western blotting. Several monoclonal antibodies were used that were specific to different epitopes on FH, shown in Figure 3-10. This would allow the differentiation of the wild type and hybrid FH species. OX24 binds to CCP 5, which was expected to be present in both FH species. Genetic testing showed that the patient was heterozygous for the Y402H polymorphism, so antibodies specific to histidine or tyrosine at amino acid position 402 (found in CCP 7) were used. The rationale was that the two FH protein species could be differentiated by the amino acid at this position. Finally L20/3 is an antibody specific for CCP 20 and therefore predicted to bind to the wild type FH protein only. The predicted hybrid protein structure, shown in Figure 3-9, did not contain the last 3 C-terminal CCPs and therefore L20/3 was not expected to bind.

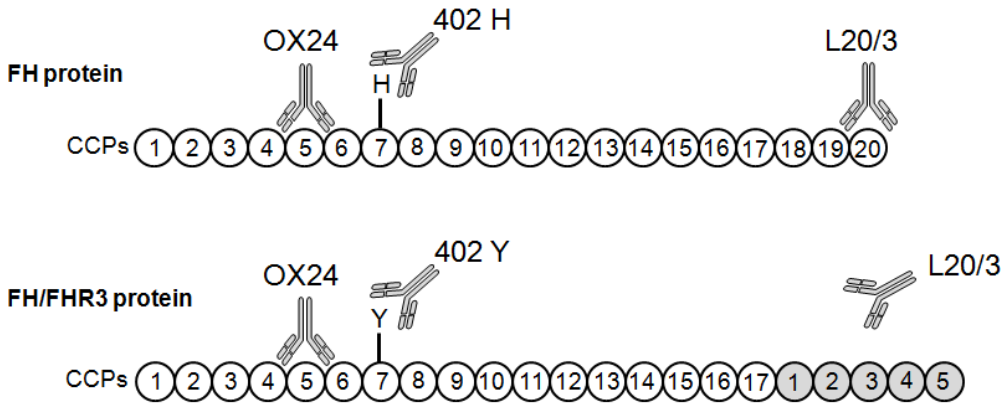

Figure 3-10 Binding sites for anti-FH antibodies.

Top- protein structure for FH. Bottom- predicted protein structure for FH/FHR3. CCPs derived from FHR3 are grey, CCPs derived from FH are white. Binding sites for anti-FH monoclonal antibodies OX24, 402H/Y and L20/3 are shown. Reprinted with permission from Challis et al. (2015).

The results of western blotting with these antibodies are shown in Figure 3-11. Figure 3-11A shows the proteins detected when the OX24 antibody was used and the wild type FH was observed in both serum samples, with a size of 150kDa. The patient serum also contained two additional bands, one running at approximately 160kDa and one at 120kDa. The upper band was hypothesised to be the hybrid FH/FHR3 and the lower band, a hybrid degradation product. When the L20/3 antibody was used, Figure 3-11B, the upper and lower bands were no longer visible in the patient sample. This indicated that the proteins lacked CCP 20, supporting the predicted FH/FHR3 structure. Finally the sera was tested using the antibodies specific for either the tyrosine or histidine at amino acid 402, shown in Figure 3-11C. When the histidine-specific antibody was used, the wild type FH was seen in the control and the patient. When the serum was probed using the tyrosine-specific antibody, the FH/FHR3 protein and its degradation product were seen in the patient sample. This confirmed that the protein in the lower band originated from the FH/FHR3 protein. The presence of the intact hybrid protein as a doublet, was hypothesised to be the result of differential glycosylation of the FHR3 portion of the protein (Skerka *et al.*, 1993).

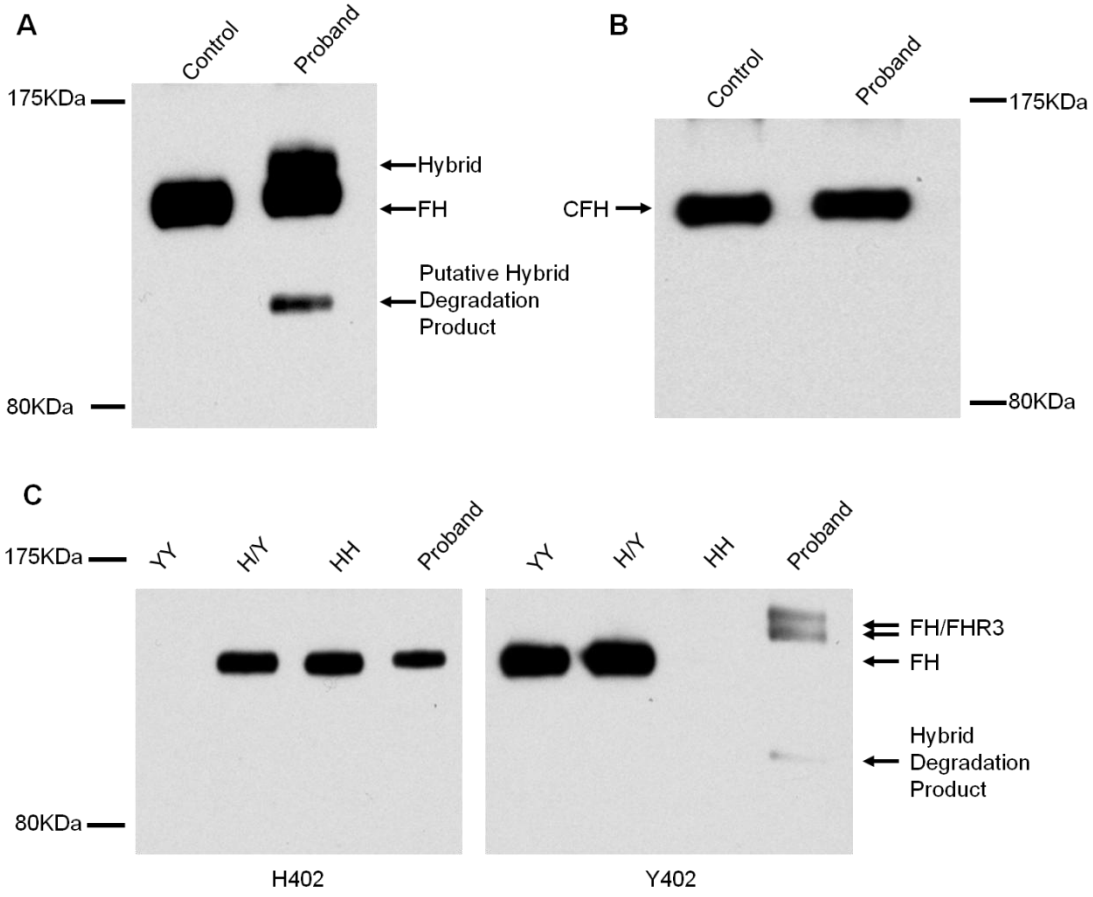

Figure 3-11 Western blotting results of sera from proband and control.

A: OX24 antibody was used. B: L20/3 antibody was used. C: Antibodies specific to H402 or Y402. Arrows indicate protein bands containing FH, FH/FHR3 and a hybrid degradation product. Reprinted with permission from Challis et al. (2015).

3.2.3.2. Mass Spectrometry

Wild type FH and FH/FHR3 species were separated from the patient serum using immunoaffinity chromatography. This was done using OX24, which was demonstrated in Figure 3-11A to be specific to both proteins. Diluted serum was loaded onto the column and unwanted serum proteins were removed during wash steps. The proteins of interest, FH and FH/FHR3, were then eluted using a low pH buffer to disrupt the bonds between OX24 on the column and the proteins. The eluted proteins were then concentrated and run on an SDS PAGE, shown in Figure 3-12.

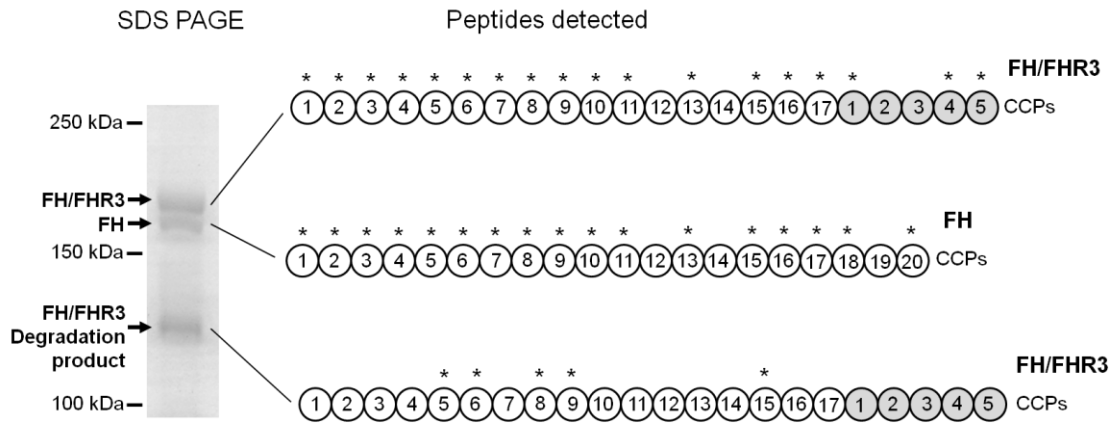

Figure 3-12 Mass spectrometry results.

Left- SDS PAGE of the isolated FH mixture from proband. Right-protein structures for predicted FH/FHR3 and FH. CCPs in grey are derived from FHR3. CCPs in white are derived from FH. '' indicate where the detected peptide fragments are identical to reference sequence of FH or FH/FHR3. Reprinted with permission from Challis et al. (2015).*

The three bands seen on the SDS page were cut out and analysed by mass spectrometry, performed by Achim Truemann. The largest band contained peptides that were consistent with the predicted FH/FHR3 protein. The intermediate band contained peptide sequences that matched the FH reference sequence. Finally the smallest band, proposed earlier to be a degradation product of FH/FHR3, contained peptide fragments that matched to FH. These 7 peptides covered 8% of FH and did not include the 402 amino acid position. As a result it did not offer further support to the earlier western blot findings (Figure 3-11C), which confirmed the 120kDa band was a degradation product of the hybrid FH/FHR3 protein. Detailed peptide sequence coverage from this experiment can be seen in Appendix F.

3.2.3.3. Cell surface decay haemolytic assay

The hybrid FH/FHR3 species was shown to lack the C-terminal region of FH, Figure 3-11B, a region that is critical for cell surface binding (Ferreira *et al.*, 2009). It was therefore predicted to have defective cell surface binding and thus, not able to regulate complement activity on the surface of cells. To test this hypothesis a cell surface decay haemolytic assay was performed. This assay measures the ability of FH species to accelerate the decay of the AP convertase on the cell surface. It was hypothesised that more OX24-purified FH and FH/FHR3 from the patient, would be required to offer the same degree of protection from lysis, as when using OX24-purified FH from a control. Figure 3-13 shows the results of the assay, performed with the help of Edwin Wong.

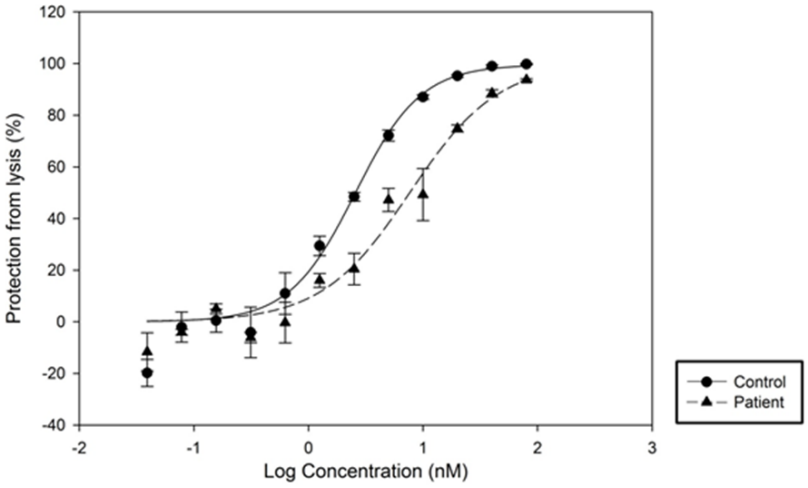

Figure 3-13 Results of the cell surface decay haemolytic assay.

Control serum- filled circles and a solid line. Patient serum- filled triangles and a broken line. X-axis shows the Log concentration of purified FH or FH and FH/FHR3 mixture used, in nM. Y-axis shows the percentage protection from lysis.

This graph demonstrated that when purified FH species from patient were used, the curve shifted to the right. This corresponded to the increased amount of protein needed to achieve comparable protection from lysis, as when using protein from the control. This supported the hypothesis that the purified FH mixture from the patient, was unable to perform as well as the control, due to the presence of half levels of wild type FH. The hybrid FH/FHR3 failed to function, due to the lack of the C-terminal domains of FH. This result was similar to what was seen by Francis *et al.* (2012) for another FH/FHR3 hybrid protein, which contained CCPs 1-19 from FH and all 5 CCPs of FHR3.

3.3. Overview of pathogenic changes in *CFH*

CFH is the most commonly mutated gene, occurring in approximately 30% (Kavanagh *et al.*, 2013, Maga *et al.*, 2010, Fremeaux-Bacchi *et al.*, 2013), of all aHUS patients. Gene defects include missense and nonsense sequence variants, genomic rearrangements and gene conversion events. This section examined all cases in the Newcastle aHUS cohort, both sporadic and familial, with *CFH* gene defects; firstly, to determine the types and frequency of genomic events in *CFH* and *CFHRs*; secondly, to review all rare point mutations and INDELs observed in *CFH*.

To date there have been five different gene rearrangements reported (Maga *et al.*, 2011, Venables *et al.*, 2006, Valoti *et al.*, 2015, Eyler *et al.*, 2013, Francis *et al.*, 2012) and three gene conversions (S1191L, V1197A or S1191L/V1197A) (Heinen *et al.*, 2006),

shown in Figure 3-14. In this project one novel gene rearrangement was found between *CFH* and *CFHR3* and one novel gene conversion in *CFHR1* was reported. The novel *CFH/CFHR3* was described in section 3.2 and labelled 'H' in Figure 3-14. The conversion of CCP20 of FH with CCP5 of FHR1, was not reported in the literature and led to the occurrence of two missense sequence variants p.L290S and p.A296V, see Figure 3-14B.

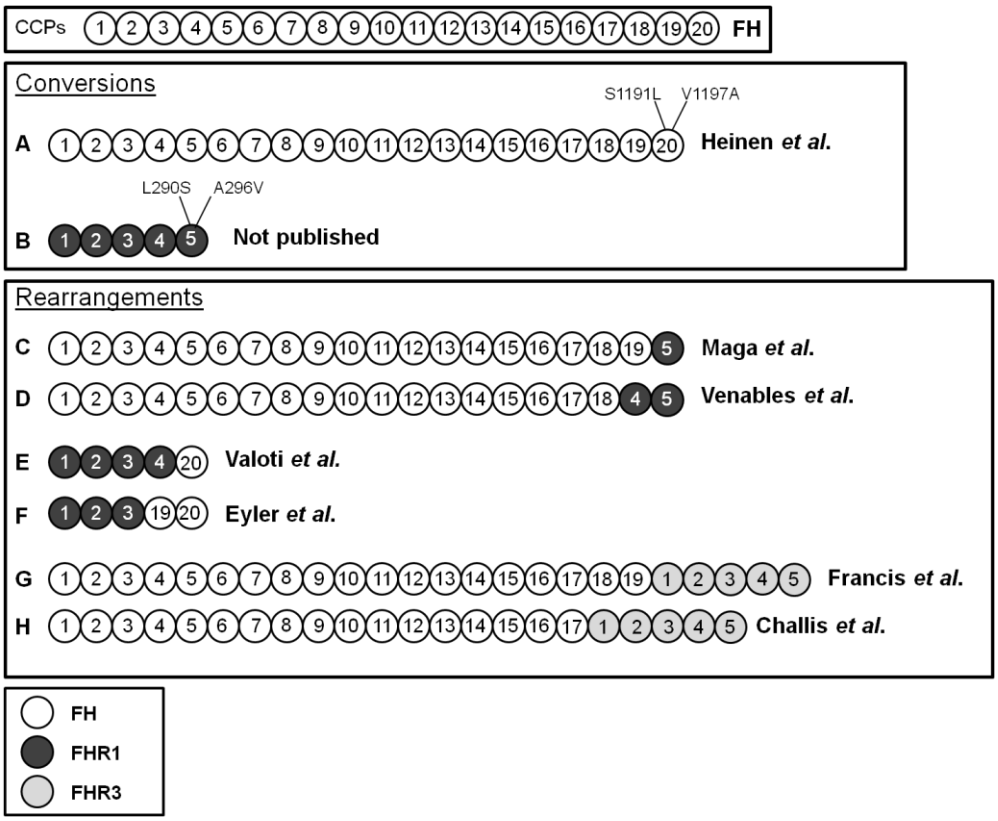

Figure 3-14 Gene conversions and rearrangements between FH, FHR1 and FHR3.

White CCPs= derived from FH, black CCPs= derived from FHR1 and grey CCPs= derived from FHR3. Structure of FH. Structure of 2 conversion proteins, A (Heinen et al., 2006) and B (not reported in literature). Structure of 6 rearrangement proteins, C (Maga et al., 2011), D (Venables et al., 2006), E (Valoti et al., 2015), F (Eyler et al., 2013), G (Francis et al., 2012) and H (described in section 3.2).

Overall 36.8% of *CFH* defects seen in the Newcastle cohort were gene rearrangements or conversions. This highlights the importance of CNV screening in aHUS patients, because Sanger sequencing does not detect gene rearrangements. The frequency of gene conversions and rearrangements in sporadic compared to familial cases was 32.5% and 60% respectively, shown in Table 9. It was interesting to observe that there were a higher proportion of gene rearrangements and conversions in familial compared to

sporadic cases. This might be because these genetic defects have increased disease penetrance.

Gene conversions A and B were identified by Sanger sequencing, to reveal two possible missense variants, S1191L and V1197A or L290S and A296V respectively. These variants can occur together or alone, depending on the extent of the gene conversion tract. Gene conversions were identified in 15.8% of the cohort; mainly in subtype A (14.5% of cohort). S1191L alone was most commonly observed, followed by V1197A and S1191L/V1197A together. Only one gene conversion type B was observed, with the presence of both missense variants, L290S and A296V. No patients had the sequence variant Q1076E, which was reported as a potential gene rearrangement between FHR1 CCP 3 and FH CCP18 (Heinen *et al.*, 2006).

Gene rearrangements were seen at higher frequency than gene conversions reported in 21.1% of the cohort. *CFHR1/CFH* reverse hybrids, type E and F, were the most frequently observed (11.6% of cohort), followed by *CFH/CFHR1* hybrids C and D (7.4%) and *CFH/CFHR3* hybrids G and H (2.1%). *CFH* and *CFHR3* are less similar, which might explain why rearrangements between them were less frequent compared to *CFH* and *CFHR1*. Rearrangements type D and F, which involved CCPs 19 and 20 of FH or CCPs 4 and 5 of FHR1, were more common than rearrangements C and E, which involved CCPs 20 of FH and CCP5 of FHR1. This was predicted to be because CCPs 19 of FH and CCP 4 of FHR1 shared the greatest sequence identity (Jozsi and Zipfel, 2008).

It was interesting to observe that MMEJ only occurred in *CFH/CFHR3* hybrids. This might be due to lower sequence homology between the two genes, meaning that HR occurs less frequently. The exact conditions that decide which DSB repair mechanism will take place, is still relatively unknown (Kass and Jasin, 2010). Ultimately the resulting proteins shared similar cell-surface binding defect as FH/FHR1 hybrids.

CFH defect			Number of cases			Percentage (%)			
Mechanism	**Subtype**		**Sporadic**	**Familial**	**Total**				
Conversion	A	S1191L	3	4	7	7.4	14.7	15.8	36.8
		V1197A	4	0	4	4.2			
		S1191L/V1197A	3	0	3	3.2			
	B	L290S	0	0	0	0.0	1.1		
		A296V	0	0	0	0.0			
		L290S/A296V	1	0	1	1.1			
Rearrangement	C		1	0	1	1.1	7.4	21.1	
	D		4	2	6	6.3			
	E		3	1	4	4.2	11.6		
	F		6	1	7	7.4			
	G		0	1	1	1.1	2.1		
	H		1	0	1	1.1			
Rare missense, splice site, nonsense, INDEL			54	6	60	63.2			
Total			**80**	**15**	**95**	**100**			

Table 9 CFH gene defects and the number of cases they are reported in.

Sporadic cases are counted as one case. Affected family members are combined to give 1 familial case (15 families in total).

The remaining 63.2% of genetic abnormalities in *CFH* were rare missense, nonsense, small INDELs or affected splice sites. These variants were plotted on the FH protein, shown in Figure 3-15. In total 44 different sequence variants were observed in the aHUS cohort, of which 41% were located in CCPs 19-20, a region that was critical for cell surface regulation (Ferreira *et al.*, 2009). Identification of the variants found within CCPs 1-18 revealed that less than half (42.3%) were missense variants, 26.9% were nonsense, 15.4% altered splice sites and 15.4% were INDELs. This suggested that variants in this area of the protein were likely to have a substantial effect on the protein. Indeed 4 out of the 11 missense variants reported, involved the loss of a cysteine residue, which was predicted to be highly detrimental to protein structure. Only 3 sequence variants were present in homozygosity, supporting previous reports that *CFH* haploinsufficiency caused aHUS (Warwicker *et al.*, 1998).

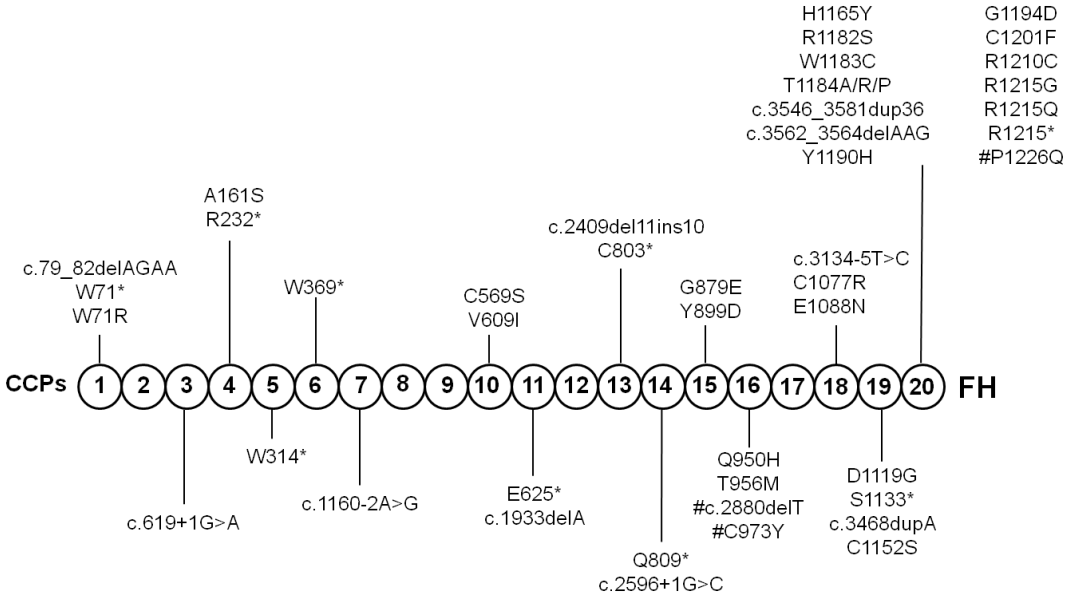

Figure 3-15 CFH mutations seen in Newcastle aHUS cohort.

All missense, nonsense, INDELs and splice site variants are plotted on FH. This information includes 2 variants described in chapter 4 (E625 and C1152S). Sequence variants that were present in homozygosity are represented by '#'.*

3.4. Discussion

The high degree of sequence homology between *CFH* and the *CFHRs* could lead to gene conversions or rearrangements leading to changes in copy number. Sanger sequencing does not detect gene rearrangements and as a result techniques such as MLPA and western blotting, were essential in detecting abnormalities.

The patient described in section 3.2 was shown to have a novel *CFH/CFHR3* gene. MLPA screening indicated the deletion of exon 23 of FH and further genetic analysis revealed that this deletion extended to the last three exons of *CFH* (exons 21-23), the result of MMEJ. This caused irregular gene splicing, leading to the incorporation of exons 2-6 of the adjacent gene, *CFHR3*. Analysis of cDNA from the patient revealed that an mRNA transcript was produced from the hybrid gene. Proteomic testing confirmed that a hybrid FH/FHR3 protein was produced in patient serum. Western blotting confirmed the loss of the C-terminal CCP20 of FH in the hybrid protein. Mass spectrometry of FH/FHR3 protein indicated the hybrid protein contained peptides identical to the FHR3 reference sequence.

Functional analysis demonstrated that the loss of the C-terminal domains of FH prevented the regulation of complement at the cell surface, similar to what was seen in another FH/FHR3 protein (Francis *et al.*, 2012). These results emphasised the importance of FH-mediated complement regulation at the cell surface, as had been previously reported (Ferreira *et al.*, 2009). This patient had onset of disease after an *E. coli* and viral infections, indicating that an environmental trigger was necessary for onset of aHUS. This mutation caused haploinsufficiency where one functional copy *CFH* was unable to prevent disease onset.

The Newcastle cohort was then analysed for all rare *CFH* sequence variants, genomic rearrangements and conversion events. Genomic rearrangements and conversion events occurred in 36.8% of cases with a *CFH* genetic aberration, of which 34.7% were between *CFH* and *CFHR1*. This was hypothesised to be the result of high sequence homology between these two genes. Gene rearrangements occurred mainly by NAHR, with the exception of two cases, which occurred by MMEJ. The exact mechanisms that determine which DNA repair mechanism is used, is not currently known. However it was hypothesised that *CFH* and *CFHR3* do not possess sufficient sequence homology for HR to take place (Rubnitz and Subramani, 1984).

Characterisation of rare sequence variants in *CFH* observed in the aHUS cohort revealed that 41% were located in CCPs 19-20, a region known to be critical for function. Variants occurring in CCPs 1-18 were shown to be highly deleterious, with 57.7% comprising of nonsense variants, INDELs or variants affecting splice sites. Only 3.2% of sequence variants (including gene rearrangements and conversions) were found

in homozygosity, indicating that *CFH*- associated aHUS has predominantly an autosomal dominant inheritance. The screening of the 28 familial aHUS cases did not identify CNVs in *CFH* or *CFHRs*, results shown in the Supplementary data. Therefore Sanger sequencing and WES were used to look for sequence variants occurring in genes known to cause TMA, described in chapter 4.

Chapter 4: Identification of mutations in known TMA-causing genes

4.1. Introduction

TMAs present with a combination of pathologies, including MAHA, thrombocytopenia and microvascular thrombosis due to endothelial injury (Scully and Goodship, 2014). Patients can experience an acute deterioration in kidney function, which requires urgent treatment (Barbour *et al.*, 2012).

One example of a TMA is aHUS, which as discussed in section 1.1.2, is the result of complement deregulation, caused by genetic variants in complement genes or autoantibodies against complement regulators (Kavanagh and Goodship, 2010). The second is Thrombotic thrombocytopenic purpura (TTP), where patients have a deficiency of A Disintegrin and Metalloproteinase with Thrombospondin type 1 motifs, member 13 (ADAMTS13) (Crawley and Scully, 2013). This is either the result of autoantibodies or inactivating mutations in *ADAMTS13* (Shenkman and Einav, 2014). ADAMTS13 is a metalloprotease that can cleave von Willebrand factor (vWF) multimers (Furlan *et al.*, 1996). vWF multimers are prothrombotic, as they promote platelet aggregation, therefore reduced ADAMTS13 activity leads to a prothrombotic environment (Sadler, 1991). Finally genetic abnormalities in *MMACHC* can lead to cobalamin c disease, where patients are at an increased risk of developing TMAs (Sharma *et al.*, 2007).

4.1.1. Aims

This chapter investigated the genetic cause of disease in familial aHUS cases with an unknown genetic aetiology. The rationale for looking only at familial cases was that it was more probable that the disease was the result of a hereditary genetic factor. Patients were screened using Sanger sequencing and WES for known TMA-causing genes *CFH*, *CFI*, *MCP*, *CFB*, *C3*, *THBD*, *MMACHC* and *ADAMTS13*.

4.2. Patient screening

In total 28 familial aHUS cases with an unknown genetic aetiology were included in this project. All families had at least one family member screened for CNVs in *CFH* and *CFHRs*, *MCP* and *FI*, by MLPA and western blotting. However no CNVs were detected, all results are shown in the Supplementary data. All available clinical data is

shown in Appendix G (Table 41) and the screening results can be seen in Appendix H (Table 42). Routine Sanger sequencing was performed for *CFH*, *MCP*, *CFI*, *C3*, *CFB* and *THBD* if sufficient DNA was available for Sanger sequencing and WES. Out of the total of 28 families, 26 were also submitted for WES. WES data was screened for sequence variants occurring in *CFH*, *MCP*, *CFI*, *C3*, *CFB*, *THBD*, *MMACHC* and *ADAMTS13*.

4.3. Factor H

Factor H (FH) is an important regulator of the alternative complement pathway, with cofactor and decay acceleration activities. It is the most commonly mutated gene in aHUS patients, as discussed previously in section 1.3.1. *CFH* sequence variants were identified in two families, described below.

4.3.1. Family 23

4.3.1.1. Clinical history

This was a family from Brazil, with two affected offspring who have since been lost in clinical follow up. DNA was available for testing for II:1 only. Routine diagnostic screening of *CFH* exons 18-23, by the Northern Genetics Service, revealed no disease-causing candidate variants.

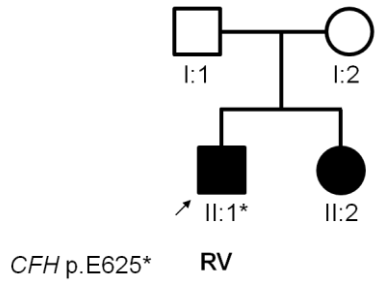

Figure 4-1 Pedigree for family 23.

Proband was indicated by the arrow. '' indicated sample was sent for WES.*

There was a limited quantity of DNA available for the proband, so screening of the remaining disease-associated genes by Sanger sequencing was not feasible. To overcome this issue, WES was carried out.

4.3.1.2. Identified sequence variant

WES revealed a heterozygous nonsense mutation in *CFH*; c.1873G>T, p.E625*. This nonsense variant was found in exon 12 of *CFH*, encoding CCP 11. This sequence variant was present on dbSNP (rs150694809) and had a MAF of 0% in 1000g and a MAF of 0.02% in ESP6500. This variant has been previously reported by Maga *et al.* (2010) in the American aHUS population.

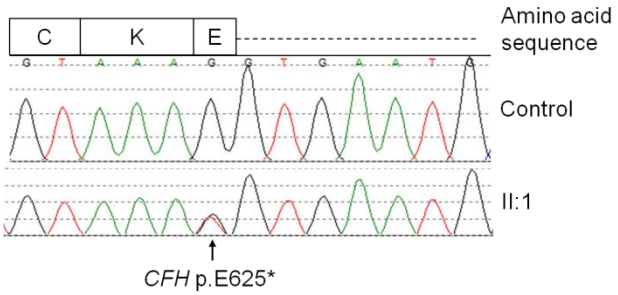

Figure 4-2 Sequencing chromatogram for CFH E625*.

The chromatograms show the sequencing results for the patient (II:1) and a control. The amino acid sequence encoded by exon 12 is shown above, where the dotted line denotes the intronic region.

4.3.1.3. In silico analysis

The gain of stop codons are severely detrimental, as it would cause the cessation of mRNA translation, leading to protein truncation or nonsense-mediated mRNA decay (Miller and Pearce, 2014). This position was highly conserved with GERP++ and PhyloP scores of 5.12 and 2.876 respectively, suggesting that it had an important function. This was supported by the 'deleterious' prediction when using Mutation taster. If the protein was secreted, it would lack the C-terminal domains, which are critical for cell surface binding (Ferreira *et al.*, 2009). This sequence variant had already been observed in conjunction with aHUS, although FH levels were not measured in this study so it cannot be established whether this nonsense variant abrogates protein secretion (Maga *et al.*, 2010).

4.3.2. Family 28

4.3.2.1. Clinical history

The proband was a 32-year old Caucasian female patient who was referred with a fever and headache, which had lasted for over a week. Laboratory evaluation showed MAHA,

thrombocytopenia, proteinuria, haematuria and elevated serum creatinine. Stool culture failed to demonstrate the presence of Shiga toxin-producing *E. coli*. ADAMTS13 activity was within the normal range (73%) and anti-FH and anti-ADAMTS13 autoantibody screening was negative. Serum levels of C4 was normal (0.28g/L), whereas the levels of C3, 80mg/dL (normal range 85-200mg/dL) and FH, 0.32g/l (0.35-0.59g/L), were low. This led to a diagnosis of aHUS and plasma exchange was initiated. The proband was then treated with eculizumab and after 20 months, had a normal creatinine and no evidence of MAHA. The proband's mother (I:2) died in her early thirties with a clinical picture consistent with a TMA.

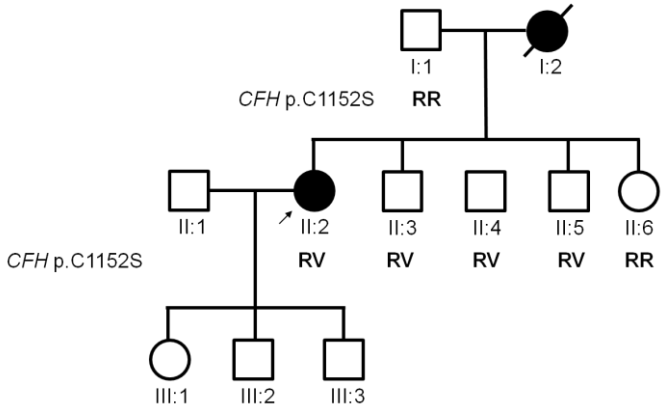

Figure 4-3 Pedigree for family 28.

Proband was indicated by the arrow. Genotype was shown in bold.

4.3.2.2. Identified sequence variant

Sanger sequencing performed by the Northern Genetics Service identified a heterozygous change in exon 21 of *CFH*, c.3454T>A, p.C1152S. This sequence variant has not been previously reported in dbSNP, ESP6500 or 1000g. It was hypothesised that it was inherited from the affected mother (I:2), although DNA was not available for testing. The father (I:1) and unaffected sister (II:6) did not carry the *CFH* variant, whereas the unaffected brothers (II:3, II:4 and II:5) did. Screening of the proband's offspring (III:1-3) was not carried out.

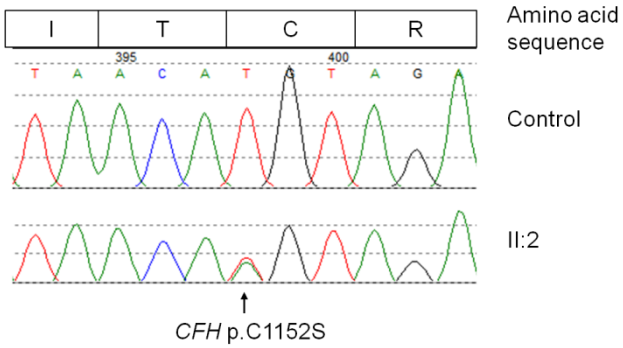

Figure 4-4 Sequencing chromatogram for CFH C1152S.

The chromatograms show the sequencing results for the patient (II:2) and a control. The amino acid sequence encoded by exon 21 is shown above.

4.3.2.3. In silico analysis

Loss of a cysteine residue is often highly detrimental to the secondary structure of a protein as it is involved in the formation of internal disulphide bridges. This was particularly important as FH is composed of 20 CCP domains, which contain 4 cysteine residues allowing the domains to be tightly compacted (Reid and Day, 1989). Figure 4-5 demonstrated the disulphide bridge formed in CCP19 between C1152 and C1109. The substitution of cysteine for a serine would disrupt the disulphide bridge and as a result, predicted to cause the protein to misfold.

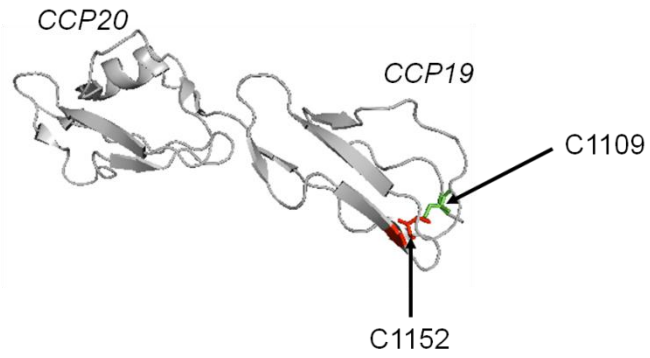

Figure 4-5 Protein model of FH C1152

Cartoon representation of factor H CCP19-20 (Protein Data Bank ID 3SWO) (Morgan et al., 2012) displayed in PyMOL. CCPs were shown in grey, red and green indicated the intra-modular disulphide bonds of C1152 and C1109, respectively.

This prediction was supported by the use of *in silico* tools, which identified C1152S as being deleterious to the protein (PolyPhen-2 HDIV and HVAR, Mutation Assessor and FATHMM), although Mutation Taster classed this as a polymorphism. This sequence variant occurred in an evolutionary conserved area, with a PhyloP score of 2.737. When

amino acid sequences were aligned in multiple species, it was found to be conserved across all species, apart from Zebrafish.

Human	N	K	R	I	T	C	R	N	G	Q	W
Mutant	N	K	R	I	T	**S**	R	N	G	Q	W
Chimp	N	K	R	I	T	C	R	N	G	Q	W
Orangutan	N	K	R	I	T	C	R	N	G	Q	W
Mouse	**K**	K	**T**	I	T	C	R	N	G	**K**	W
Rat	N	K	**I**	**V**	T	C	R	N	G	**K**	W
Rabbit	**D**	K	**K**	I	**M**	C	R	N	G	**K**	W
Dolphin	**K**	**R**	**H**	I	**V**	C	**Q**	N	G	**E**	W
Dog	N	**R**	**N**	I	**I**	C	**S**	N	G	Q	W
Opossum	**S**	**R**	**F**	**V**	T	C	**K**	N	G	**R**	W
Platypus	**S**	K	**D**	**V**	**R**	C	**E**	N	G	**K**	W
Zebrafish	-	-	-	-	-	-	-	-	-	-	-

Figure 4-6 Amino acid sequence alignment for FH C1152.

Human wild type and mutant sequences are compared to 10 species. Grey shaded column indicated position of C1152. Amino acids in bold indicate a difference to the human reference sequence.

These deleterious predictions were supported by the low levels of FH observed in the proband, suggesting that misfolding of the protein prevented protein secretion. It was interesting to observe that the siblings (II:3, II:4 and II:5) who also carried this sequence variant, did not have disease. This coincided with previous findings that disease penetrance was not complete and that additional factors might be necessary to cause disease (Kavanagh *et al.*, 2013).

4.4. Factor I

FI is a serine protease that can inactivate C3b and C4b by removing their alpha chains, in the presence of a cofactor, such as MCP and FH (Kavanagh *et al.*, 2005). Mutations in *CFI* have been associated with aHUS, as discussed in section 1.3.3. A *CFI* sequence variant was identified in a familial case of aHUS, described below.

4.4.1. Family 27

4.4.1.1. Clinical history

This was a family with two affected daughters, both of whom are now deceased. The proband was a 16 year old female, who presented with headache and vomiting. Further analysis indicated acute renal failure with hypertension, thrombocytopenia and MAHA. She then suffered a cardiac arrest and died. Diagnostic screening revealed normal serum levels of FH (0.45g/L) and C3 (1.33g/L). Her older sister (II:1) died at 19 years of age

after a similar episode, following recurrent aHUS-like episodes over the preceding 12 months. The mother died at 39 years of age, although her cause of death was unclear.

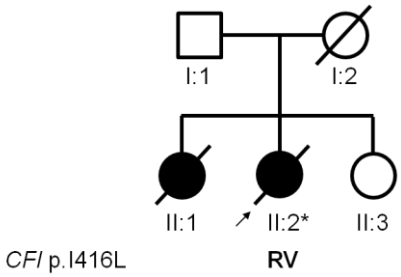

Figure 4-7 Pedigree of family 27.

Proband was indicated by the arrow. '' indicated sample was sent for WES. Heterozygous genotype was indicated by RV.*

4.4.1.2. Identification of Genetic variants

WES led to the discovery of a heterozygous *CFI* mutation in the proband, c.1246A>C, p.I416L. This was listed on dbSNP (rs61733901) and had a MAF of 0.14% in 1000g database and 0.4% in ESP6500. This change was found in exon 12 and encodes part of the serine protease domain of FI.

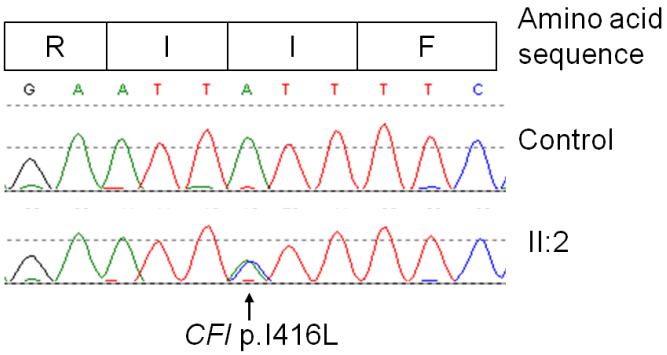

Figure 4-8 Sequencing chromatogram for CFI I416L.

The chromatograms show the sequencing results for the patient (II:2) and a control. The amino acid sequence encoded by exon 12 is shown above.

4.4.1.3. In silico analysis

The pathogenicity of *CFI* I416L was predicted using *in silico* tools. GERP++ and PhyloP predicted this variant to be in an evolutionary conserved position, with scores of 5.62 and 2.145 respectively. If the amino acid sequence of FI was aligned at the position of I416 and compared to other species, it was observed that the isoleucine was conserved in all species apart from rat.

Human	Y	N	E	H	F	I	I	R	D	V	Y
Mutant	Y	N	E	H	F	**L**	I	R	D	V	Y
Chimp	Y	N	E	H	F	I	I	R	D	V	**H**
Orangutan	Y	N	E	H	F	I	I	R	D	V	Y
Mouse	Y	**K**	E	H	**V**	I	**V**	R	**K**	V	**T**
Rat	-	-	-	-	-	-	-	-	-	-	-
Rabbit	Y	N	E	H	**I**	I	I	**K**	**N**	V	**L**
Dolphin	Y	**K**	**D**	H	**I**	I	I	**Q**	**N**	V	**L**
Dog	Y	N	E	H	**V**	I	I	**Q**	**K**	**A**	**L**
Opossum	Y	**D**	E	H	**I**	I	**V**	**Q**	**K**	V	**R**
Platypus	Y	**Q**	**D**	H	**V**	I	**V**	**Q**	**N**	V	**R**
Zebrafish	Y	N	E	H	**I**	I	I	**K**	**K**	V	-

Figure 4-9 Amino acid sequence alignment of FI I416.

Human wild type and mutant sequences are compared to 10 species. Grey shaded column indicated position of I416. Amino acids in bold indicate a difference to the human reference sequence.

These results suggested that this amino acid may be important in FI function. Indeed I416L was predicted to be deleterious by all *in silico* tools apart from PolyPhen-2 and Mutation Assessor, which classed it as benign and medium impact, respectively. Protein modelling of I416L on FI was carried out to determine a possible mechanism for disease, shown in Figure 4-10. It was observed that I416L was located in close

proximity to the catalytic triad H362, D411 and S507 (Morley and Walport, 2000), suggesting that it may affect FI serine protease activity.

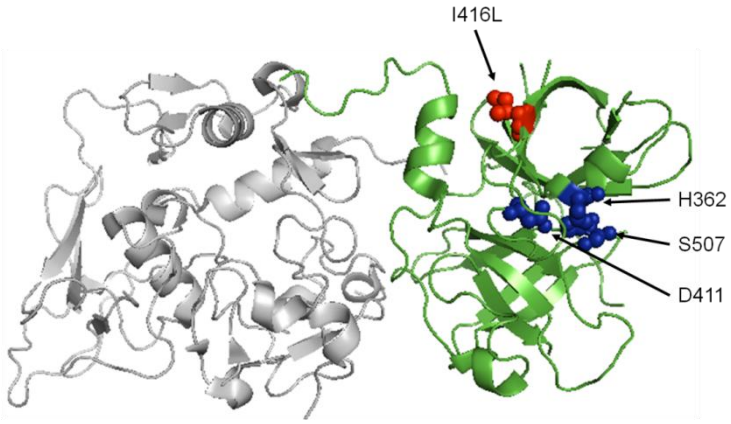

Figure 4-10 Protein model of FI I416L.

FI structure obtained from Protein Data Bank (2XRC) (Roversi et al., 2011) and viewed in PyMOL. FI is shown in a cartoon format, with the heavy chain in grey and the serine protease domain in green. Amino acids that make up the catalytic triad, H362, D411 and S507 (Morley and Walport, 2000), were shown as blue spheres. I416L was shown by red spheres.

4.4.1.4. Evidence in literature

This sequence variant had been previously reported by Sellier-Leclerc *et al.* (2007) in the French aHUS cohort, (annotated as p.I398L when numbered from beginning of mature protein). It was found in a patient with low C3 (348mg/L, compared to a normal range of 660-1260mg/L) and FI (70% of the expected amount) levels in the serum. Bienaime *et al.* (2009) then found 3 additional sporadic aHUS cases with I416L. Of these three individuals, 2 had low C3 levels and out of two with FI levels tested, 1 had low FI levels. Finally I416L had also been reported in one patient in the American aHUS cohort, although no clinical information was provided (Maga *et al.*, 2010). Serum levels of FI for the proband were not available, however it was hypothesised that they were low, coinciding with observations of Bienaime *et al.* (2009) and Sellier-Leclerc *et al.* (2007).

Bienaime *et al.* (2009) then performed *in vitro* functional analysis by recombinantly transfecting HEK293 cells with a wild type or mutant FI vector. They demonstrated that cells with the I416L mutant had lower FI levels in supernatant than cells containing the wild type FI vector. It was proposed that low FI secretion was due to the protein being retained within the cell. To test this hypothesis immunofluorescence was carried out on the mutant and control HEK293 cells. Control FI protein was found to colocalise with

wheat germ agglutinin, a marker of the trans-Golgi network, whereas the I416L mutants colocalised with protein disulfide isomerise, a biomarker of the endoplasmic reticulum (Bienaime *et al.*, 2009).

4.5. Membrane Cofactor Protein (MCP)

MCP (Membrane cofactor protein) is a cell surface bound complement regulator, which acts as a cofactor to FI in the cleavage of 3b and C4b (Seya *et al.*, 1986). *MCP* mutations have been reported in association with aHUS, as described in section 1.3.4. This section described *MCP* sequence variants identified in 4 familial aHUS cases.

4.5.1. Family 4

4.5.1.1. Clinical history

This family had two affected family members, across two generations (father and son). I:1 had normal complement levels (FH-0.67g/L, FI-47mg/L), 2 copies of *CFHR1-3* and was negative for FH-autoantibodies. The proband had normal complement levels (FH-0.61g/L, FI- 46mg/L), 2 copies of *CFHR1-3* and was negative for anti-FH autoantibodies. Routine Sanger sequencing of *CFH* revealed no mutations.

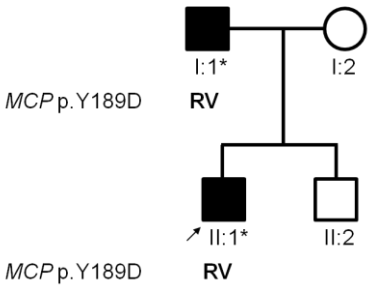

Figure 4-11 Pedigree for family 4.

Proband was indicated by the arrow. '' indicated sample was sent for WES.*

4.5.1.2. Identified sequence variant

WES revealed a heterozygous mutation in exon 5 of *MCP*, c.565T>G, p.Y189D that segregated with disease. It was listed on dbSNP as rs202071781, had a MAF of 0.02% on 1000g and was not seen in ESP6500. This variant had been previously reported in aHUS patients (Fremeaux-Bacchi *et al.*, 2006, Sullivan *et al.*, 2010, Maga *et al.*, 2010).

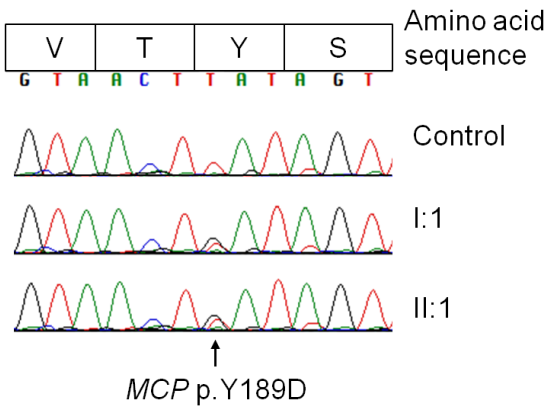

Figure 4-12 Sequencing chromatogram for MCP Y189D.

Top was the result for the control, then I:1 and at the bottom was II:1. Y189D is indicated by an arrow. The amino acid sequence encoded by exon 5 is shown above.

4.5.1.3. In silico analysis

In silico analysis indicated that Y189 was a conserved amino acid, with a GERP++ score of 4.85 and a PhyloP score of 2.158. Figure 4-13 shows the amino acid sequence alignment of Y189 in 10 species. The tyrosine residue was highly conserved in all species except for platypus and zebrafish.

Human	L	D	A	V	T	Y	S	C	D	P	A
Mutant	L	D	A	V	T	**D**	S	C	D	P	A
Chimp	L	D	A	V	T	Y	S	C	D	P	A
Orangutan	L	**E**	A	V	T	Y	S	C	D	P	A
Mouse	**H**	**E**	A	V	**S**	Y	S	C	D	P	**T**
Rat	**H**	**E**	A	V	**I**	Y	S	C	D	P	**N**
Rabbit	**A**	**E**	**S**	V	T	Y	S	C	D	P	**S**
Dolphin	**N**	**E**	**L**	V	T	Y	S	C	**N**	P	**S**
Dog	**N**	**E**	**V**	V	T	Y	S	C	**E**	**K**	**S**
Opossum	**G**	**E**	**T**	V	T	Y	**H**	C	-	-	**L**
Platypus	-	-	-	-	-	-	-	-	-	-	-
Zebrafish	-	-	-	-	-	-	-	-	-	-	-

Figure 4-13 Amino acid sequence alignment of MCP Y189.

Human wild type and mutant sequences are compared to 10 species. Grey shaded column indicated position of Y189. Amino acids in bold indicate a difference to the human reference sequence.

Y189D was predicted to have a deleterious effect on the protein by PolyPhen-2 (HDIV and HVAR), Mutation Assessor and RadialSVM. Mutation Taster and FATHMM reported it to be 'neutral' and 'tolerated' respectively. To determine a possible mechanism for disease, Y189D was plotted on a protein model of CCPs 1-4 of MCP. Y189D was found in CCP3 close to the hinge region between CCPs 3 and 4, an area

involved in C3b and C4b binding (Liszewski *et al.*, 2000). This might indicate that Y189D could alter ligand binding, reducing its cofactor activity.

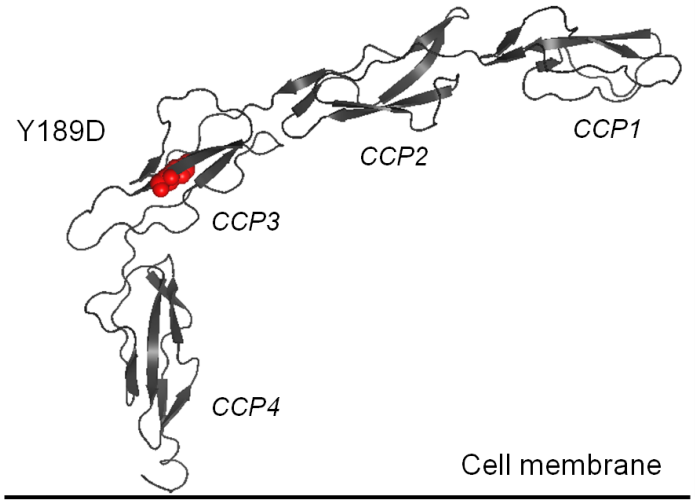

Figure 4-14 Protein model of MCP Y189D.

Predicted protein structure of MCP CCPs 1-4 (amino acids 35-285), created using Phyre2. CCPs are shown as cartoons in grey. Red spheres indicate the position of Y189D. Protein structure was positioned against the cell membrane, adapted from (Persson et al., 2010).

4.5.1.4. Evidence in literature

Experiments *in vitro* demonstrated that Y189D might cause a structural abnormality, leading to retention of protein in the endoplasmic reticulum. This was shown by the increase in the ratio of precursor and mature MCP observed, in cells transfected with a mutant vector (Fremeaux-Bacchi *et al.*, 2006). This led to the hypothesis that Y189D altered protein folding, which caused a reduction in protein expression on the cell surfaces. Fremeaux-Bacchi *et al.* (2006) reported three aHUS patients with Y189D, who also carried either D185N on the same allele or IVS7-2A>G on an alternate allele. They found that patient peripheral blood mononuclear cells (PBMCs) containing the mutant *MCP* gene, had reduced cell surface expression, supporting *in vitro* experiments. Reduction of MCP expression would cause inadequate complement regulation on the surface of the cells, leading to onset of disease. Samples from the proband and the proband's father were not available for testing, so reduced surface expression of MCP could not be demonstrated.

4.5.2. Families 18 and 22

4.5.2.1. Clinical history

Family 18 had an affected sibling pair. DNA was available for testing from both affected individuals. The proband's FH serum level (0.70g/L) was normal and genetic screening of *CFH* exons 19-23 revealed no candidate mutations. II:2 had normal FH (0.68g/L) and FI (63mg/L) levels, was negative for anti-FH autoantibodies and genetic screening of *CFH* exons 19-23 revealed no mutations. MLPA screening revealed 2 copies of *CFHR1/3* in both II:1 and II:2.

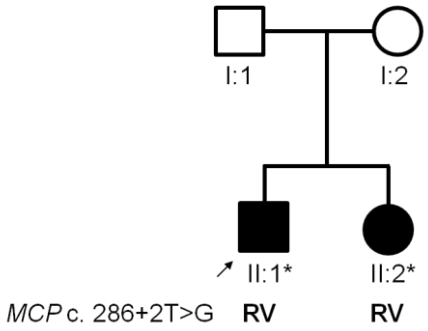

Figure 4-15 Pedigree for family 18.

Proband was indicated by the arrow. '' indicated sample was sent for WES. Genotype was shown in bold.*

Family 22 contained two affected brothers. The proband had normal serum levels of complement proteins C3 (0.75g/L), C4 (0.20g/L), FH (0.50g/L) and FI (49mg/L). MLPA indicated the patient had 1 copy of *CFHR1/3*. Diagnostic screening of complement genes (*C3, CFH and CFI*) failed to identify any rare sequence variants.

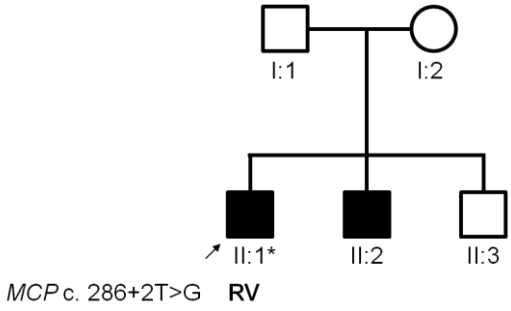

Figure 4-16 Pedigree for family 22.

Proband was indicated by the arrow. '' indicated sample was sent for WES. Genotype was shown in bold.*

4.5.2.2. Identified sequence variant

The probands in families 18 and 22 were both found to have a sequence variant in *MCP*, c.286+2T>G. The affected sibling in family 18 (II:2) was also demonstrated to carry this variant in heterozygosity, shown in Figure 4-17. This variant occurred in the 5' end of intron between exons 2 and 3, causing disruption to the splice donor site. This variant, also referred to as IVS2+2T>G, had been previously reported in the French (Fremeaux-Bacchi *et al.*, 2006), American (Maga *et al.*, 2010) and Dutch (Westra *et al.*, 2010) aHUS cohorts. It was listed on dbSNP (rs186757635), reported in 1000g with a MAF of 0.02% and was not found on ESP6500.

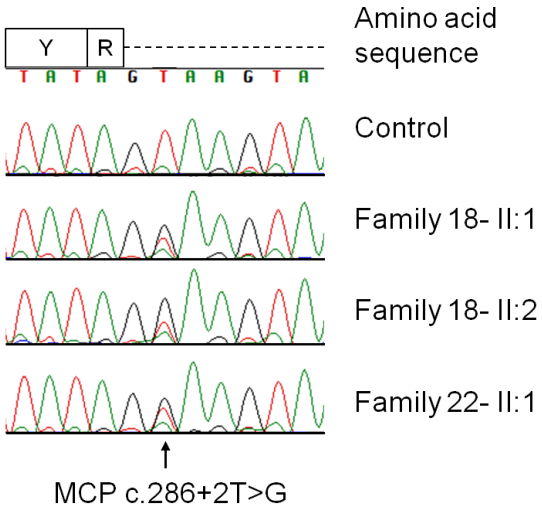

Figure 4-17 Sequencing chromatogram for MCP c.286+2T>G.

The chromatograms show the sequencing results for the patients and a control. The amino acid sequence encoded by exon 2 is shown above, the dotted line denotes the intronic region.

4.5.2.3. In silico analysis

In silico analysis found that the nucleotide substitution would reduce the score of the donor splice site from 79.30 (maximum was 100) to 62.18, suggesting that it might have caused abnormal gene splicing. This variant was predicted to be deleterious by Mutation Taster and conserved by GERP++ (2.33) and PhyloP (0.661).

4.5.2.4. Evidence in literature

Fremeaux- Bacchi *et al.* (2006) sequenced mRNA from a patient who was homozygous for c.286+2T>G. They identified that this sequence variant caused the 45th nucleotide of exon 2 to splice into exon 3, causing a 144bp deletion from the 3' end of exon 2. The

resultant protein would have lost 48 amino acids. This patient was shown to have between 0-25% MCP expression on granulocytes compared to controls. It was predicted that the patients described here, who were heterozygous for c.286+2T>G would have approximately 50% reduction in the surface expression of MCP. However no patient material was available for testing to corroborate this hypothesis.

4.5.3. Family 26

4.5.3.1. Clinical history

Family 26 was referred to the National aHUS Service from Saudi Arabia. The parents were not cousins but were from the same tribe and therefore classed as consanguineous. The parents (I:1 and I:2) and two unaffected offspring II:1 and II:5 had no reported renal phenotype. There had been 4 affected offspring, one of whom had died.

II:2 first presented at 12 years old (now 24 years old) required transfusions to mediate frequent occurrences of thrombocytopenia. Individual II:3 was 15 years old when he had severe thrombocytopenia. He did not respond well to plasma therapy and died. II:4 presented at 14 years old with frequent attacks of thrombocytopenia. He had one severe attack which was complicated with a central nervous system haemorrhage, however currently had normal renal function. The proband (II:6) presented at 10 years of age with flu-like symptoms, fever, thrombocytopenia and elevated serum creatinine. Serum C3 levels were borderline low (0.7g/L, normal range 0.68-1.38 g/L) and C4 levels were normal.

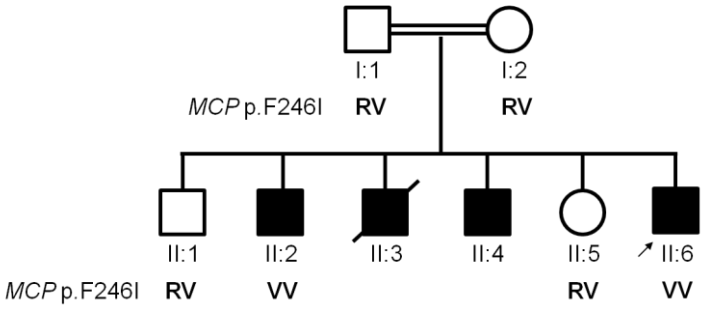

Figure 4-18 Pedigree for family 26.

Parents were consanguineous, indicated by double line. Proband was indicated by the arrow. MCP F246I genotype was indicated by 'RV' for heterozygous or 'VV' in homozygous.

4.5.3.2. Genetic variant

Routine diagnostic testing of genes associated to aHUS, performed by the Northern Genetics Service, revealed the sequence variant c.736T>A, p.F246I in *MCP*. This was found in exon 4, encoding CCP 4 of MCP. The proband (II:6) and one of the other affected offspring (II:2), were identified as homozygous for the variant (no DNA was available for testing for II:3 and II:4). The parents and 2 unaffected offspring (II:1 and II:5) were shown to carry the variant in heterozygosity. This sequence variant was not previously described in any aHUS cohort, nor was it observed in 1000g or ESP6500 databases.

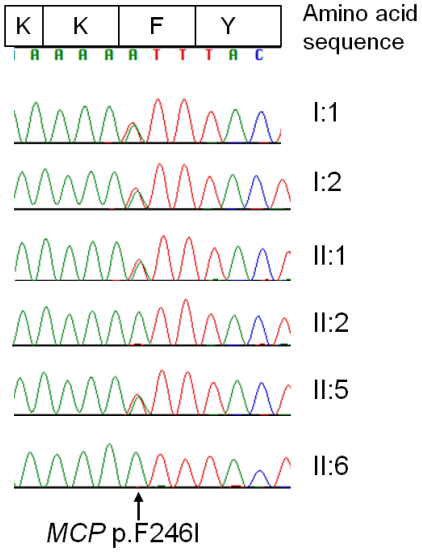

Figure 4-19 Sequencing chromatogram for MCP F246I.

Each chromatogram refers to a family member from the pedigree shown in Figure 4-18. The amino acid sequence encoded by exon 4 is shown above. All affected individuals tested (II:2 and II:6) were homozygous for the change, whereas all remaining family members were heterozygous.

4.5.3.3. In silico analysis

In silico analysis was used to determine whether F246I occurred in an area that was conserved across evolution. Firstly PhyloP gave this sequence variant a score of 0.729, indicating it was evolutionary conserved. Secondly sequence alignment against 10 other species, shown in Figure 4-20, indicated that this amino acid was conserved in half of the species tested.

Human	G	F	G	K	K	F	Y	Y	K	A	T
Mutant	G	F	G	K	K	**I**	Y	Y	K	A	T
Chimp	G	F	G	K	K	F	Y	Y	K	A	T
Orangutan	G	F	G	K	K	F	Y	Y	K	A	T
Mouse	G	**A**	G	**E**	**I**	F	**S**	Y	**Q**	**S**	T
Rat	**R**	**T**	**E**	K	K	F	**S**	-	-	-	-
Rabbit	G	F	**R**	K	K	**Y**	Y	Y	**Q**	A	T
Dolphin	G	**S**	**R**	K	K	F	**S**	Y	**Q**	A	**V**
Dog	G	F	**S**	**R**	K	**Y**	Y	Y	K	A	**R**
Opossum	G	F	G	**S**	**S**	**H**	**T**	Y	K	**D**	T
Platypus	-	-	-	-	-	-	-	-	-	-	-
Zebrafish	-	-	-	-	-	-	-	-	-	-	-

Figure 4-20 Amino acid sequence alignment of MCP F246.

Human wild type and mutant sequences are compared to 10 species. Grey shaded column indicated position of F246. Amino acids in bold indicate a difference to the human reference sequence.

F246I was then analysed to see if it was predicted to cause a detrimental effect to the MCP's structure and function. PolyPhen-2 HDIV and HVAR classed this variant as deleterious, whereas Mutation Taster, Mutation Assessor and FATHMM classed it as 'polymorphism', 'medium impact' and 'tolerated', respectively. This might indicate that this missense substitution had a lesser impact on the protein and only when present in homozygosity, would it cause disease. In order to investigate the potential effect of F246I on protein function, it was plotted on the protein model of CCPs 1-4 of MCP. F246I was positioned in CCP4, close to the hinge region between CCPs 3 and 4 and therefore predicted to alter protein conformation.

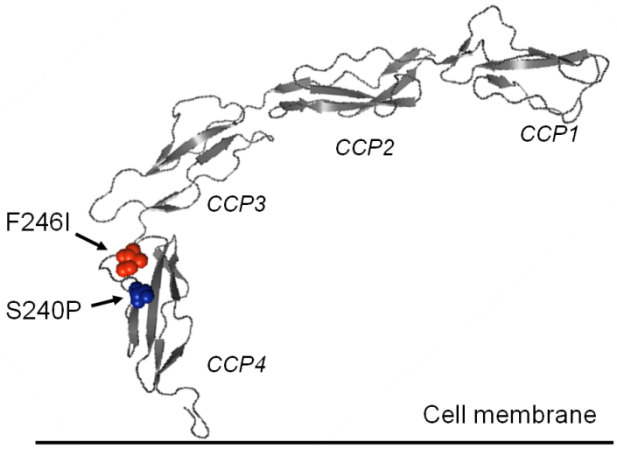

Figure 4-21 Protein model of MCP F246I.

Predicted protein structure of MCP CCPs 1-4, created using Phyre2. CCPs are shown as cartoons in grey. Red spheres indicate the position of F246I and blue spheres indicate S240P. Protein structure was positioned against the cell membrane, adapted from Persson et al. (2010).

Functional work had been performed on a nearby mutation S240P (Richards *et al.*, 2003). Patients with this mutation were shown to express MCP on their PBMCs, at levels comparable to controls. However *in vitro* experiments identified that the substitution abrogated C3b binding. No patient samples were available to test MCP expression on cell surfaces. It was hypothesised that the protein was either not present at the cell surface or if it was expressed, then it would not bind to C3b and would have no cofactor activity.

The occurrence of MCP deficiency is very rare (10% of MCP-aHUS) (Liszewski *et al.*, 2007) and interestingly, nearly a half of these patients also develop common variable immunodeficiency (Fremeaux-Bacchi *et al.*, 2006, Couzi *et al.*, 2008). This may be because MCP-activated T cells induce B-cells more strongly, leading to increased Ig expression (Fuchs *et al.*, 2009).

4.6. ADAMTS13

4.6.1. Gene background

A Disintegrin and Metalloproteinase with Thrombospondin type 1 motifs, member 13 (*ADAMTS13)* contains 29 exons and is located on chromosome 9q34 (Levy *et al.*, 2001). It encodes a metalloprotease that can cleave large vWF multimers into smaller subunits (Crawley and Scully, 2013). Loss of ADAMTS13 activity leads to a build up of ultra large vWF multimers leading to a prothrombotic disease, known as TTP

(Stockschlaeder *et al.*, 2014). TTP can be caused by sequence variants in *ADAMTS13* or autoantibodies that inhibit ADAMTS13, both of which result in low ADAMTS13 activity, approximately <5-10% (Scully and Goodship, 2014, Zheng *et al.*, 2010). Sequence variants are inherited in an autosomal recessive or compound heterozygous pattern (Peyvandi *et al.*, 2004). Figure 4-22 demonstrates all sequence variants reported in *ADAMTS13* to date.

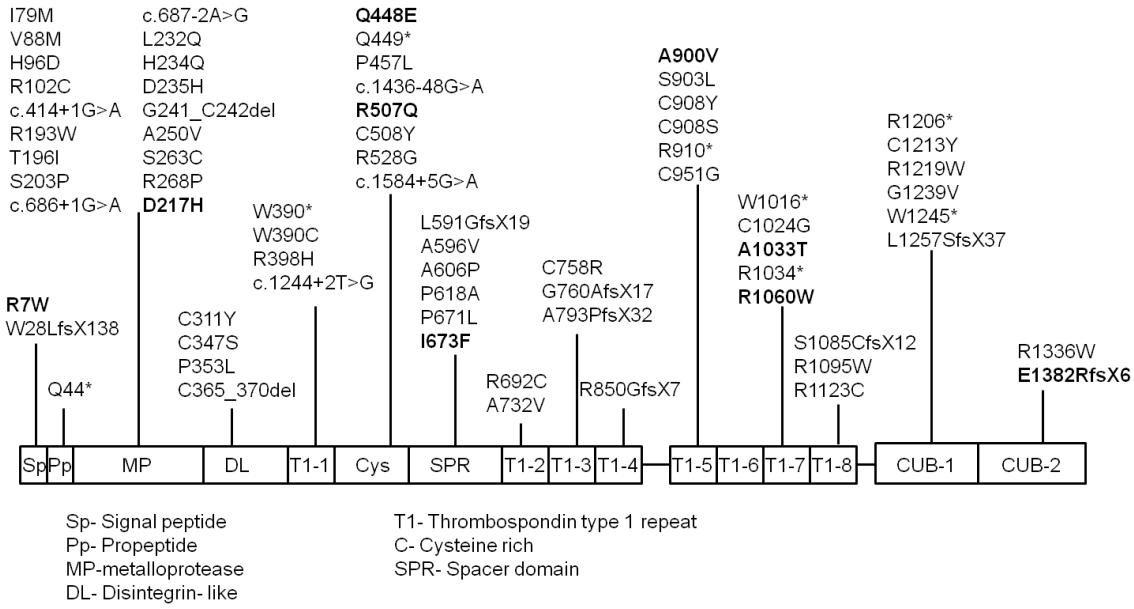

Figure 4-22 Diagram showing sequence variants reported ADAMTS13.

Variants were taken from TTP database (Peyvandi) and scientific literature, rare and common SNPs were included. The protein structure was adapted from Plaimauer et al. (2006). Variants in bold, indicated that they were identified here.

This section described two families that were found to have compound heterozygous sequence variants in *ADAMTS13*.

4.6.2. Family 7

4.6.2.1. Clinical history

Family 7 had a history of TMA in the family. The father (II:2) and daughter (III:1) were referred with a diagnosis of aHUS and the paternal uncle (II:1) had a diagnosis of TTP. The father (II:2) was screened for mutations in *CFH, MCP, CFI* and *CFB*, which were found to be normal. The proband was found to have 1 copy of *CFHR1/3* and was negative for anti-FH autoantibodies.

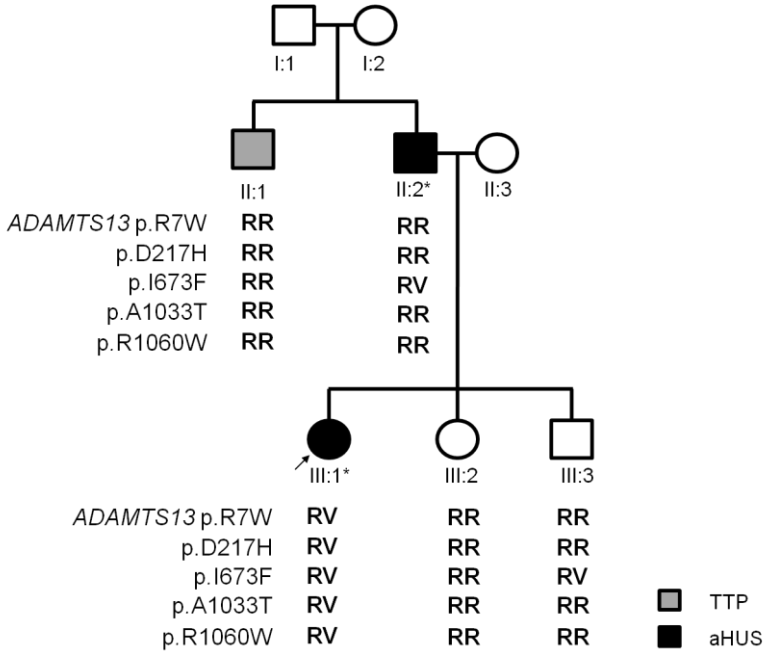

Figure 4-23 Pedigree for family 7.

Grey- indicated family members with TTP, black- with aHUS.'' indicated samples sent for WES. Proband was shown by that arrow. Heterozygous genotype was indicated by RV.*

4.6.2.2. Identification of Genetic variants

The initial analysis of known complement genes by Sanger sequencing failed to reveal a genetic cause for disease, therefore WES was undertaken. This led to the identification of several mutations in a TMA-associated gene *ADAMTS13*. These changes were then confirmed by Sanger sequencing in multiple family members, shown in Table 10. No DNA was available for the mother (II:3). R7W had the highest MAF, at 10% and 12% in ESP6500 and 1000g respectively and was reported in dbSNP database (rs34024143). A1033T was seen in ESP6500 and 1000g at 3% and reported in dbSNP (rs28503257). R1060W was reported in ESP6500 at 0.08%, 1000g at 0.05%, in dbSNP (rs142572218) and was previously reported in association with TTP (Tao *et al.*, 2006, Camilleri *et al.*, 2008). Neither D217H nor I673F were reported in ESP6500, 1000g databases or dbSNP, however both have been previously reported in association with TTP (Camilleri *et al.*, 2012, Matsumoto *et al.*, 2004).

Sequence variant		Exon	II:1	II:2	III:1	III:2	III:3	MAF (%)
Transcript	Protein							
c.19C>T	p.R7W	1	RR	RR	RV	RR	RR	12
c.649G>C	p.D217H	6	RR	RR	RV	RR	RR	0
c.2017A>T	p.I673F	17	RR	RV	RV	RR	RV	0
c.3097G>A	p.A1033T	24	RR	RR	RV	RR	RR	3
c.3178C>T	p.R1060W	24	RR	RR	RV	RR	RR	0.05

Table 10 Sequence variants identified in ADAMTS13 in family 7.

R= Reference allele, V= Variant allele. MAF was given as a percentage, based on 1000 genome database.

The results indicated that that the proband was compound heterozygous for changes in *ADAMTS13*. I673F originated from the father and the other sequence variants were hypothesised to have been inherited from the mother or were *de novo*. The absence of maternal DNA, prevented this from being determined. It was apparent that the paternal uncle (II:1) and father (II:2) were not compound heterozygous for sequence variants in *ADAMTS13*, suggesting that neither was affected by TTP. Deeper clinical phenotyping was performed and demonstrated that both the father and uncle were not affected and the proband had evidence of haemolysis, normal renal function and a confirmed ADAMTS13 activity of 2.5%. Therefore this was in fact a sporadic case of TTP; however this case will continue to be referred to as family 7. The pedigree was redrawn to demonstrate this, shown in Figure 4-24.

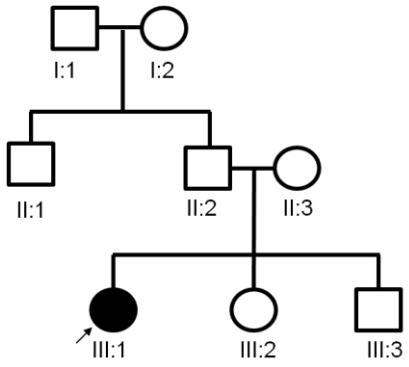

Figure 4-24 Updated pedigree for family 7.

4.6.2.3. Functional significance

In silico analysis was carried out to determine if the sequence variants occurred in conserved regions and if they were predicted to be deleterious. GERP++ and PhyloP were used to investigate whether variants occurred in an evolutionary conserved area. Sequence variants D217H, A1033T and R1060W had scores above the preset thresholds

of 2 and 0.5 for GERP++ and PhyloP, respectively. I673F had a score above zero but below the conservation threshold, indicating it was semi-conserved. R7W was identified as not conserved, with negative scores in both programs.

Sequence variant	GERP++	PhyloP
R7W	-6.55	-1.237
D217H	4.69	2.161
I673F	1.6	0.027
A1033T	5.59	2.64
R1060W	3.17	1.192

Table 11 Conservation scores for ADAMTS13 variants in family 7.

The amino acid sequence of ADAMTS13 was then aligned with 10 other species, at the position of the 5 missense substitutions, shown in Figure 4-25. The findings were comparable to the conservation scores produced using *in silico* tools. R7 was not well conserved, only occurring in the reference sequence for chimp and orangutan. D217, I673, A1033 and R1060 were highly conserved, each maintained across at least 9 other species.

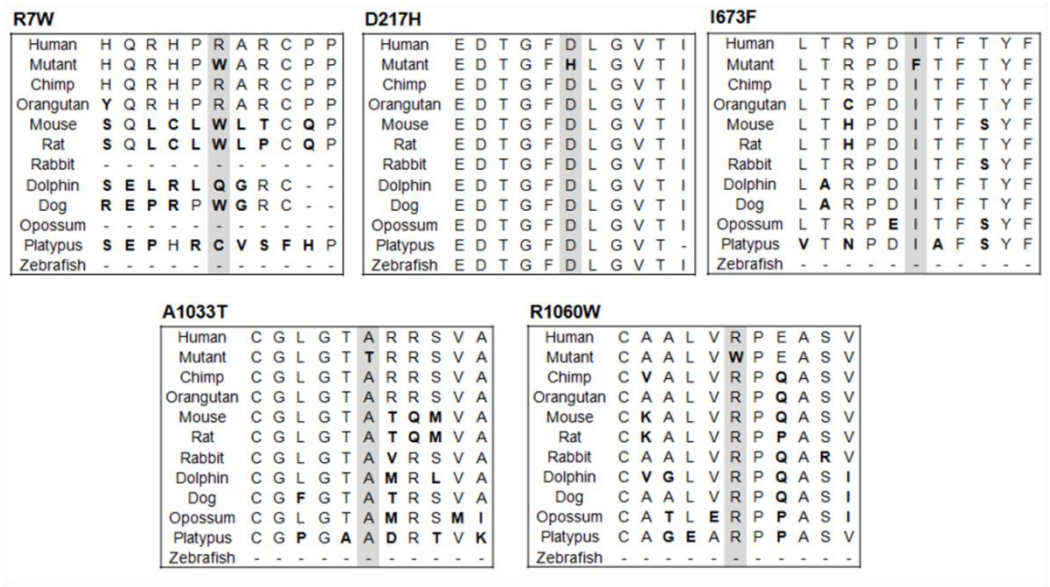

Figure 4-25 Amino acid sequence alignment for ADAMTS13 variants in family 7.

Human wild type and mutant sequences are compared to 10 species. Grey shaded column indicated position of amino acid substitution. Amino acids in bold indicate a difference to the human reference sequence.

The functional impact of these sequence variants was then investigated, shown in Table 12. D217H was described as deleterious by all prediction programs, with the exception of Mutation Assessor, which classed it as a low impact substitution. I673F, A1033T and

R1060W were predicted to be deleterious by PolyPhen-2 HDIV and HVAR and Mutation Taster, yet considered to be tolerated by FATHMM and RadialSVM. Mutation assessor predicted I673F and R1060W to have a medium impact and A1033T to have a low impact on the protein. Finally R7W was predicted to be deleterious by FATHMM and either benign, tolerated or a polymorphism by the remaining programs.

Variant	*In silico* testing					
	PolyPhen-2		Mutation Taster	Mutation Assessor	FATHMM	RadialSVM
	HDIV	HVAR				
R7W	B	B	P	N	D	T
D217H	D	D	D	L	D	D
I673F	D	D	D	M	T	T
A1033T	D	D	D	L	T	T
R1060W	D	D	D	M	T	T

Table 12 In silico results of family 7 ADAMTS13 sequence variants.

B= benign, D= deleterious, L= low impact, M= medium impact, N= neutral, P= polymorphism and T= tolerated.

4.6.2.4. Evidence in literature

In vitro analysis had been performed on several *ADAMTS13* mutations, which identified defects in terms of protein secretion and activity, shown in Table 13. Feng *et al.* (2013) found that secretion of protein was comparable to the control in R7W-transfected cells, although there was a slight decrease in protein activity (86%). I673F caused proteins to be retained within transfected HeLa cells (Matsumoto *et al.*, 2004). D217H was shown to be secreted in transfected HEK293T cells but had greatly reduced protein activity (Camilleri *et al.*, 2012). A1033T was shown to be secreted normally in transfected HEK293T cells, but had slightly decreased activity (80%) (Feng *et al.*, 2013). Although in HeLa cells the activity of A1033T was shown to be comparable to wild type (Tao *et al.*, 2006). Camilleri *et al.* (2008) demonstrated using cell lysates that R1060W did not affect metalloprotease activity. However it led to reduced protein expression to 5% of normal in HEK293T cells (Camilleri *et al.*, 2008) and 11% of normal in HeLa cells (Tao *et al.*, 2006).

Sequence variant	dbSNP ID	Domain	*In vitro* testing	
			Secretion (%)	Activity (%)
R7W	rs34024143	Sp	99*	86*
D217H	-	MP	112[+]	24[+]
I673F	-	SPR	0**	ND
A1033T	rs28503257	T1-7	100*^	80*, 100^
R1060W	rs142572218	T1-7	5[#], 11^	100[#]

Table 13 In vitro results of ADAMTS13 sequence variants seen family 7.

*Protein activity and secretion are given as a percentage compared to wild type ADAMTS13. Sp= Signal peptide, MP= Metalloprotease, SPR= Spacer, T1-7= Thrombospondin Type 1-repeat 7, ND= not done, *=(Feng et al., 2013), **=(Matsumoto et al., 2004), [+]= (Camilleri et al., 2012), [#]=(Camilleri et al., 2008) and ^=(Tao et al., 2006).*

R1060W and I673F have been demonstrated using *in vitro* assays, to lead to reduced protein secretion and D217H reduced ADAMTS13 activity. Genotyping of the father and the unaffected siblings demonstrated that I673F was likely to have been inherited from the father. The absence of maternal DNA meant that it was unknown if D217H and R1060W were inherited together from the mother or if one had occurred *de novo*. The combination of these variants on alternate alleles would cause defective ADAMTS13 activity, which was demonstrated in the proband (2.5% activity).

4.6.3. Family 24

The pedigree for Family 24 was shown in Figure 4-26. They were referred to the National aHUS Service as having aHUS. The proband had normal levels of FH (0.61g/L), 1 copy of *CFHR1/3* and was negative for anti-FH autoantibodies. Genetic testing of *CFH* exons 18-23 revealed no pathogenic sequence variants. The affected sister was also shown to have 1 copy of *CFHR1/3* and no anti-FH autoantibodies. She underwent genetic testing of *CFI*, which revealed no candidate sequence variants.

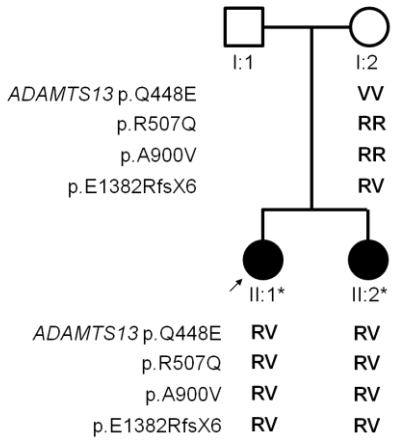

Figure 4-26 Pedigree of family 24.

Proband was indicated by the arrow. '' indicated sample was sent for WES. Genotype was shown in bold.*

4.6.3.1. Identification of Genetic variants

WES identified multiple sequence variants in *ADAMTS13*, which segregated with disease. These variants were confirmed by Sanger sequencing in the affected siblings and the mother, shown in Table 14. DNA for the father was not available. R507Q was absent from ESP6500 and 1000g databases, although it had been previously reported by Veyradier *et al.* (2004) in a patient with TTP. E1382RfsX6 was reported in ESP6500, in 1 out of 12,515 alleles and previously reported in association with TTP (Kentouche *et al.*, 2002, Pimanda *et al.*, 2004). Q448E and A900V were seen more frequently, with MAFs in 1000g of 27% and 7.7%, respectively and in ESP6500 at 30.1% and 11.3%, respectively.

Sequence variant		dbSNP ID	Exon	I:2	II:1	II:2	MAF (%)
Transcript	Protein						
c.1342C>G	p.Q448E	rs2301612	12	VV	RV	RV	27
c.1520G>A	p.R507Q	rs281875296	13	RR	RV	RV	0
c.2699C>T	p.A900V	rs685523	21	RR	RV	RV	7.7
c.4142_4143insA	p.E1382RfsX6	rs387906343	29	RV	RV	RV	0

Table 14 Sequence variants identified in ADAMTS13 in family 24.

R= Reference allele, V= Variant allele. MAF was given as a percentage based on 1000g database.

4.6.3.2. In silico analysis

In order to predict if the sequence variants would be pathogenic, several *in silico* tools were used. The sequence conservation was examined using GERP++ and PhyloP, shown in Table 15. Q448E, R507Q and A900V were classed as evolutionary conserved

by GERP++ and PhyloP. The conservation score for E1382RfsX6 could not be calculated using either program.

Sequence variant	GERP++	PhyloP
Q448E	4.25	0.678
R507Q	5.42	2.535
A900V	4.55	0.971

Table 15 Conservation scores for ADAMTS13 variants in family 24.

Scores could not be calculated for E1382RfsX6, because INDELs are not compatible with either program.

The amino acid sequence was then aligned with 10 other species. Q448 and A900 were poorly conserved, which differed from the predictions of GERP++ and PhyloP. R507 was shown to be highly conserved, apart from in Zebrafish. Finally E1382 was maintained across 6 species, indicating that it was also evolutionary conserved.

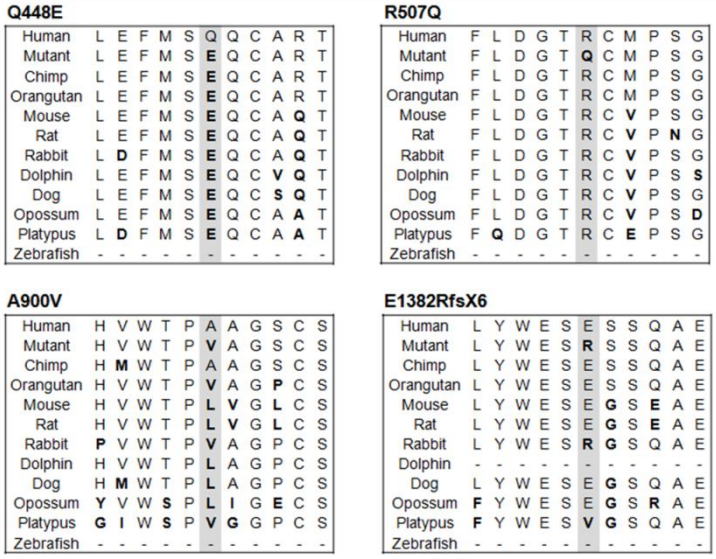

Figure 4-27 Amino acid sequence alignment for ADAMTS13 variants in family 24.

Human wild type and mutant sequences are compared to 10 species. Grey shaded column indicated position of amino acid that is changed by a mutation. Amino acids in bold indicate a difference to the human reference sequence.

In silico tools were used to predict the effects of the variants on the protein, shown in Table 16. Predictions could not be calculated for E1382RfsX6. However due to the occurrence of a premature stop codon, it was predicted to be functionally significant. R507Q was predicted to be deleterious by PolyPhen-2 (HDIV and HVAR) and Mutation Taster. It was then classed as having a 'medium' impact on ADAMTS13 by Mutation Assessor and predicted to be tolerated by FATHMM and RadialSVM. Q448E

and A900V were both predicted to be benign by PolyPhen-2 HDIV and HVAR, polymorphisms by Mutation Taster, neutral by Mutation Assessor and tolerated by FATHMM and RadialSVM.

Variant	*In silico* testing					
	PolyPhen-2		Mutation Taster	Mutation Assessor	FATHMM	Radial SVM
	HDIV	HVAR				
Q448E	B	B	P	N	T	T
R507Q	D	D	D	M	T	T
A900V	B	B	P	N	T	T

Table 16 In silico results of ADAMTS13 sequence variants seen family 24.

B= benign, D= deleterious, N= neutral, P= polymorphism and T= tolerated. Predictions could not be calculated for E1382RfsX6.

These predictions suggested that Q448E and A900V would have little effect on the protein, whereas R507Q would be highly detrimental. E1382RfsX6 was predicted to be damaging to the protein, as it led to the formation of a premature stop codon.

4.6.3.3. Evidence in literature

Feng *et al.* (2013) carried out *in vitro* analysis of *ADAMTS13* sequence variants A900V and Q448E. They found that A900V had 0% activity and that Q448E had a slight reduction in activity (75%) and secretion (95%) that of the wild type ADAMTS13. Pimanda *et al.* (2004) demonstrated that E1382RfsX6 nearly entirely abolished protein secretion in transfected cells.

Sequence variant	Domain	*In vitro* testing	
		Secretion (%)	Activity (%)
Q448E	Cysteine rich	95**	75**
R507Q	Cysteine rich	ND	ND
A900V	T1-5	ND	0**
E1382RfsX6	CUB-2	14^{+}	85^{+}

Table 17 In vitro results of ADAMTS13 sequence variants seen family 24.

*Secretion and activity are represented as a percentage of normal activity. **(Feng et al., 2013) or $^{+}$(Pimanda et al., 2004)*

Schneppenheim *et al.* (2003) identified carriers of E1382RfsX6 had ADAMTS13 levels, less than half that seen in healthy controls. Unaffected carriers of R507Q were also shown to have half the normal level of ADAMTS13 (Veyradier *et al.*, 2004). The mother was confirmed to carry E1382RfsX6 and not R507Q, which confirmed that R507Q was on the alternate allele. This suggested that the proband and affected sibling were compound heterozygous and thus had absent ADAMTS13 activity, causing TTP.

Due to the presence of R507Q in both affected offspring, it is predicted to have been inherited from the father. It appeared unlikely that the same sequence variant arose *de novo* in two separate events in one family. However the absence of paternal DNA prevented this from being tested.

4.7. Discussion

In this project 28 familial aHUS cases with an unknown genetic cause, were examined for sequence variants occurring in genes known to be associated with TMAs. This included *CFH*, *CFI*, *MCP*, *CFB*, *C3*, *THBD*, *MMACHC* and *ADAMTS13*.

Routine Sanger sequencing identified two novel sequence variants, one in *CFH* and one in *MCP*. The proband of family 28 carried the sequence variant C1152S in *CFH*, in heterozygosity. *In silico* analysis predicted that this amino acid was highly conserved and deleterious to the protein. The loss of a cysteine was hypothesised to disrupt protein folding and as a result alter protein secretion, which was supported by the low serum levels of FH. Unaffected family members also carried this variant, indicating incomplete disease penetrance. The *MCP* F246I was observed in a consanguineous family, inherited in an autosomal recessive pattern. *In silico* predictions indicated that it was in an evolutionary conserved area and deleterious in 2 out of 5 programs. Protein modelling demonstrated that it was positioned near the hinge region between CCPs 3 and 4, suggesting it might impact protein conformation and possibly lead to retention of protein within the cell. Functional work on a nearby variant S240P, demonstrated that this region might be involved in C3b binding, so F246I could alternatively cause disease by reducing MCP cofactor activity.

The remaining 26 families were submitted for WES and of these, 7 were found to have sequence variants in genes known to be associated with TMAs (*CFH*, *MCP*, *CFI* and *ADAMTS13*). The *CFH* change E625* was very rare and *in silico* analysis determined that this was a conserved amino acid and that the gain of a stop codon would be deleterious, suggesting it was likely to be the genetic cause of disease. This sequence variant had been previously described in an aHUS patient, further supporting this hypothesis. The mutation I416L in *CFI* was rare, in an evolutionary conserved region and predicted to have a serious impact on protein structure and function. It was previously reported in patients with low FI levels in serum, indicating that this sequence variant abolished protein expression.

Two sequence variants were observed in *MCP*, Y189D and c.286+2T>G, both of which were autosomal dominant in inheritance. These variants were rare and had been previously reported in the literature in association with aHUS. They were both predicted to be conserved and deleterious. Y189D and c.286+2T>G were demonstrated in the literature to cause reduced cell surface expression of MCP. Samples from the patients described here, were not available to confirm a reduction in MCP expression on cells.

Two families were referred as having a clinical diagnosis of aHUS. Genetic screening of the affected individuals found compound heterozygous changes in *ADAMTS13*. Review of clinical data for one family (family 7) led to the observation that there was no renal involvement in the proband. Therefore these patients were reclassified as having TTP. This suggested that the similar phenotypes in aHUS and TTP could lead to misdiagnoses, which would have a serious implication in the treatment received by patients.

WES was used as a screening tool to analyse genes, which due to a lack of DNA, would not have been possible. However WES only covers approximately 1% of the genome and as a result, it is possible that some disease-causing sequence variants were missed. The sequence variants identified in these genes were very rare and predicted to be deleterious by *in silico* tools. In addition the majority of sequence variants were previously described in the literature in association with disease, with the exception of *CFH* C1152S and *MCP* F246I, which were both novel.

The screening of familial aHUS cases yielded no rare sequence variants that segregated with disease in *CFB*, although rare, functionally significant changes have been identified in the sporadic cohort. Screening of *THBD* in familial cases did not identify rare variants that segregated with disease. Further screening of 75 sporadic cases only identified two SNPs (data shown in the Supplementary data). Therefore no evidence was found in this cohort that supported *THBD* as being associated with disease, comparable to what was observed in the French aHUS cohort (Fremeaux-Bacchi *et al.*, 2013). Finally no sequence variants were identified in *MMACHC* in either the familial or sporadic cohort. This may be due to the distinctive phenotype, which means that clinicians are not referring these patients. The remaining 19 families were analysed with a multistep, candidate pipeline to try and identify novel gene candidates, described in chapter 5.

Chapter 5: Novel TMA-causing genes

5.1. Introduction

There are many approaches that can be used to identify sequence variants that cause a disease. Firstly Sanger sequencing can be carried out, although this is very time consuming and requires prior knowledge (Ku *et al.*, 2012). Linkage analysis identifies genes that segregate in individuals with disease, however genetic heterogeneity can be confounding (Bishop, 1994). Homozygosity mapping can be performed on consanguineous families, identifying autosomal recessive genes that have been inherited from parents (Lander and Botstein, 1987). Yet linkage analysis and homozygosity mapping have poor resolution, identifying large areas that contain many candidate genes. Therefore additional analysis by Sanger sequencing is required to identify the candidate sequence variant. Finally GWAS can detect common sequence variants that may have an association with disease. However these variants are more likely to have a small effect on disease, accounting for a small proportion of disease heritability (McCarthy *et al.*, 2008).

Next generation sequencing technologies such as WES and WGS have the ability to screen vast amounts of DNA. This is beneficial as it enables screening of genes without prior knowledge, allowing the diagnosis of other pathologies, which is advantageous if the phenotype is unclear or an atypical presentation. The rational for utilising WES, was that despite examining approximately 1% of the genome, the exome is thought to contain around 85% of Mendelian disease causing sequence variants (Choi *et al.*, 2009). Therefore compared to WGS, it is comparatively faster and more cost effective (Bamshad *et al.*, 2011).

WES was first successfully used to identify disease-causing sequence variants in Freeman-Sheldon syndrome (Ng *et al.*, 2009). They demonstrated that the use of filtering steps could reduce the number of candidate variants for disease. It has since been successful in the detection of novel candidate genes in many diseases, such as Miller syndrome (Ng *et al.*, 2010b) and Kabuki syndrome (Ng *et al.*, 2010a). The use of multistep filtering pipelines can help to prioritise candidate variants (Bamshad *et al.*, 2011). This can include using an appropriate mode of inheritance, frequency of the variant in the general population and predicted functional impact (Gilissen *et al.*, 2012).

However this has limitations and can lead to removal of potential candidate variants (Erlich *et al.*, 2011).

5.1.1. Chapter aims

The proportion of patients in aHUS cohorts, who have no genetic aetiology, had been previously reported to be approximately 30-50% (Sellier-Leclerc *et al.*, 2007, Maga *et al.*, 2010, Fremeaux-Bacchi *et al.*, 2013). In the Newcastle aHUS cohort 55% of families did not have a genetic cause of disease. The aim of utilising WES was to try and identify novel genetic causes of aHUS.

5.2. Patient screening

In this project 28 families with no known genetic cause for disease were analysed. Chapter 3 described the rearrangements and gene conversions that occurred in *CFH* and the *CFHRs*. In the Newcastle aHUS cohort, the frequency at which these changes arise in the patients with *CFH* abnormalities was 36.8%. Since these rearrangements are not always detected by Sanger sequencing methods, where possible, patients were also screened using MLPA (*CFH*, *CFHRs*, *CFI* and *MCP*) and western blotting (FH and FHRs). No abnormalities were detected suggesting that the genetic cause of disease was not due to CNVs occurring in these genes. However there is still the possibility that CNVs have occurred elsewhere in the genome.

The first stage of the analysis involved screening patients for sequence variants in genes associated with TMAs, shown in Table 44 of Appendix J. The results of this screening were described Chapter 4, where 9 families were found to have sequence variants in genes associated with TMAs.

The second stage of the analysis investigated the genetic cause of disease in the remaining 19 families. The probands in these families and where available, additional family members, were analysed using WES. Overall there were 6 families with 1 family member screened by WES, 11 families with 2 tested, 1 family with 4 tested and 1 family with 7 tested. In total 5 families in this study were consanguineous. 4 samples were whole genome amplified (WGA) due to insufficient quantities of DNA for WES. The use of WGA samples in WES has been previously described, some found the sequencing results were comparable to non-amplified samples (Hollegaard *et al.*, 2013,

Hasmats *et al.*, 2014). However Pinard *et al.* (2006) demonstrated that WGA samples had lower target coverage and increased bias, reducing data quality.

5.3. WES raw data

The raw data received was separated into whether or not they had been WGA prior to WES; the complete list was shown in Appendix E. The mean and standard deviation (SD) was calculated for each group, shown in Table 18, for the total number of reads received, the number of reads that mapped to reference genome (after removal of duplicate reads) and the percentage of exome covered at various read depths.

The total number of reads and the total number of reads mapped to the reference genome (after removal of duplicate reads), the mean coverage and percentage coverage (at all read depths) was higher in non-amplified samples compared to WGA samples This is similar to what has been previously reported (Rykalina *et al.*, 2014). This may be because WGA samples were not evenly amplified, and therefore during exome enrichment, the target region will not be evenly represented (Pinard *et al.*, 2006). This results in a library with low complexity, where fewer reads will map uniquely to the reference genome.

		Non WGA (n=46)		WGA (n=4)	
		Mean	SD	Mean	SD
Total number of reads received		99723087.3	35933194.4	63117915.5	16355994.0
Total number of reads mapped to the reference genome (after removal of duplicates)		91009750.0	29863736.6	60214572.0	13824960.1
Mean coverage (x)		73.0	16.7	42.2	16.0
Percentage of exome Covered	1x	97.8	1.2	82.5	23.5
	5x	95.6	3.7	68.0	31.9
	10x	92.1	7.3	56.7	30.5
	20x	82.1	12.5	38.9	24.7
	40x	56.4	16.8	19.6	14.2

Table 18 Read and coverage data for WES samples.

The total number of reads received from the sequencing provider and the total number of reads that map to the reference genome, once duplicates have been removed (arise from amplification bias). SD=Standard deviation.

WES allows for the interrogation of functionally important regions of the genome, as previously described in section 1.6. This is an unbiased approach and will lead to the generation of large quantities of data. It can be difficult to determine the sequence variants that are responsible for aHUS and those that are not. Table 19 shows the number of variants before any filtering has taken place, after exonic and splice site (2 intronic nucleotides either side of exon) are positively selected for and after an initial removal of common sequence variants (with a MAF of greater than 5% in ESP6500 and 1000g). These steps were carried out in the bioinformatic pipeline shown in Figure 2-5.

The average starting number of variants for WGA samples was 412,499 compared to 210,825 in non-WGA samples. This was not as expected as the quality of the sequencing for WGA samples was worse than non-amplified samples, shown by the reduced exome coverage at all read depths (Table 18). Therefore it was predicted that more the variants in the WGA samples would be removed during the quality control (QC) steps of the bioinformatic pipeline. The reasons for this could not be determined. There were also far higher numbers of variants in WGA samples than non-amplified samples, after removing variants outside regions of interest and with frequencies $\geq$5%. For example Families 13 and 14 have an average of 19,621 variants, compared to an

average of 2,032 variants for all other families. This was also unexpected and the reasons for this observation were not known. The variant list for family 5 also contains a large number of variants, which is the result of the inclusion of the WGA sample (II:5) and the large sample number (7 family members tested in total).

Family	Total number of variants		
	Total	Exonic and splice site	MAF <5%
1	220101	15807	1755
2	160304	12866	1320
3	215791	15517	1718
5*ǂ	418076	20252	3775
6	207680	13493	1321
8	238937	15948	1856
9	234571	15710	1824
10	321184	21243	3325
11*ǂ	331942	19031	2609
12	229649	16661	2101
13*ǂ	428821	31327	24746
14*	471155	24479	14495
15ǂ	161227	17186	1938
16	230854	17365	1924
17	142845	13849	1254
19	203247	17577	2075
20	239632	17220	2081
21	190075	17092	1893
25	166279	10891	1772

Table 19 Total number of variants observed in a family.

The total number of variants are shown for the 19 families with no known genetic aetiology. This is a culmination of all variants seen in all family members, where variants occurring in multiple family members are reported once. Firstly the total number of variants in the annotation file before any filtering has taken place. Secondly the number of variants that are located in an exon or splice site. Thirdly the number of variants with a MAF of <5% in 1000g and ESP6500. These filtering steps were carried out as part of the bioinformatic pipeline. indicates families containing WGA samples, ǂ indicates consanguineous families.*

5.4. Variant filtering strategy

Filtering strategies are a series of steps that aim to reduce the number of candidate variants to a smaller, more manageable number. This includes using tools that either look at the frequency, functional impact or conservation of a sequence variant, shown in Figure 5-1. This analysis aimed to identify the good candidates for disease, i.e. rare,

conserved and with a detrimental impact on the protein. Although it was possible that this method would lead to rejection of a causative sequence variant.

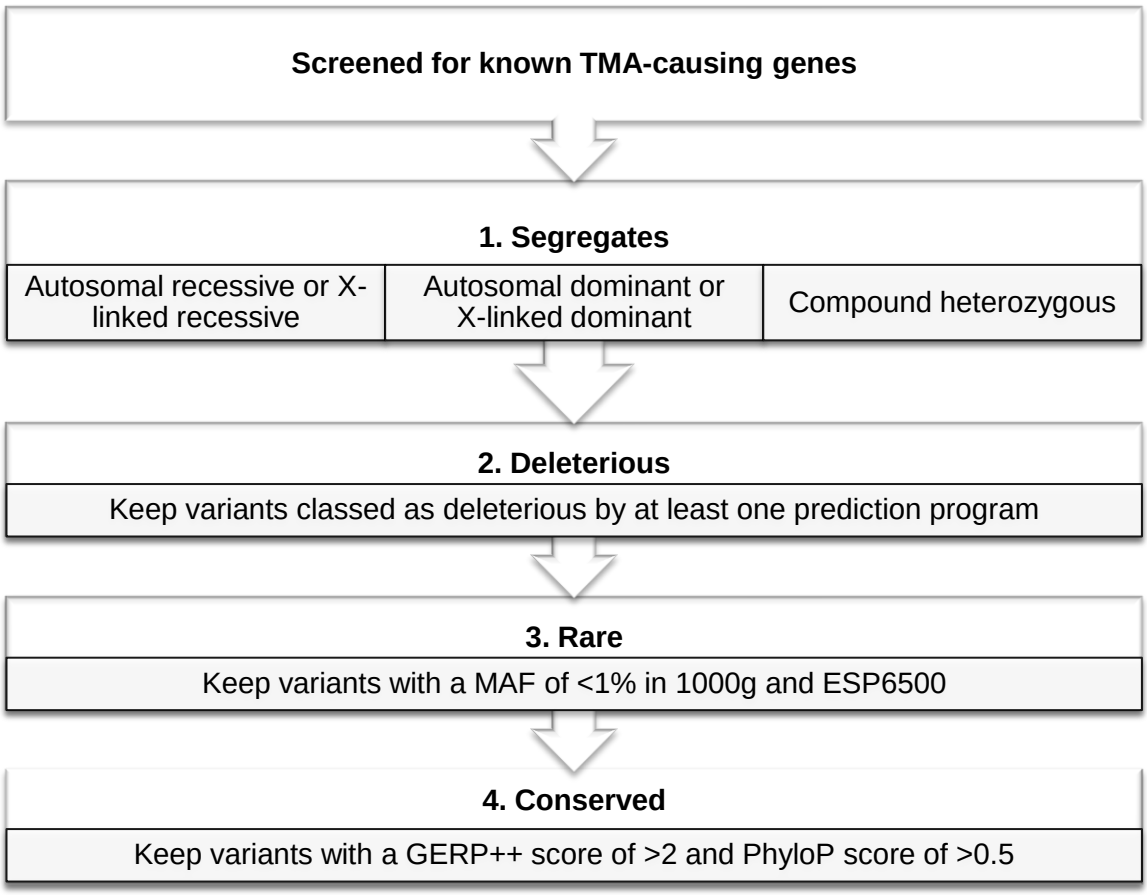

Figure 5-1 Variant filtering strategy.

This was performed for remaining 19 families with no genetic cause of disease. Each analysis was performed three times, using the three different inheritance patterns listed in step 1.

5.4.1. Variant segregation

There are several possible mechanisms for pathogenic variants to be inherited, such as recessive (homozygous and compound heterozygous variants), X-linked or autosomal dominant. Variants can also arise as *de novo* events, however due to the presence of multiple affected individuals within a family; it was assumed that variants were inherited by descent. The formatting of the data in Excel sheets meant that it was necessary to separate the analyses for homozygous and compound heterozygous variants. However they were later combined to try and identify candidate genes, where a complete functional deficiency was necessary to cause disease.

To perform these analyses, several assumptions were made. Variants inherited in a dominant fashion were considered to have complete penetrance. This meant that only affected individuals carried the sequence variant. For a recessive mode of inheritance, all affected offspring must contain the same pathogenic variants in homozygosity, with the assumption of complete penetrance. If parental samples were available for testing, both must have the same variant in heterozygosity. Any unaffected siblings with DNA available for screening must not be homozygous for the variant.

For compound heterozygous analysis it was assumed that the affected offspring inherited one defective allele from each parent. If there were affected family members in two generations, such as mother and daughter, then this analysis was contraindicated and thus not performed. Complete penetrance was assumed, where all affected offspring must have the two sequence variants and unaffected siblings must not. Variants cannot be present in homozygosity in any family member.

Finally X-linked inheritance was considered if male offspring were affected. The software used annotates these variants as VV (homozygous for variant allele), despite there being only one allele present. Therefore these variants were included in the list of variants that were autosomal recessive. Any variants with poor coverage and given a genotype 'NA' (not available) by annotation software were removed from the analysis.

These assumptions allow for stringent filtering, leading to the formation of a smaller candidate list. However there are caveats with the methods used. Firstly there is the possibility that variants have arisen *de novo* and therefore would not follow the requirements set out previously. This was considered improbable, due to the occurrence of multiple affected family members. Secondly any variants with poor coverage have been removed from the analysis, which could remove a sequence variant of interest. However inclusion of these results would increase the number of variants in the final candidate list, making it difficult to find potential disease-causing variants. Thirdly the disease penetrance has been assumed to be complete. Yet this may not be true, as other genes associated with aHUS do not have complete penetrance (Kavanagh *et al.*, 2013). However data for unaffected family members, which allows negative selection to take place, was only available in two families (families 5 and 15).

5.4.2. In silico functional predictions

The next step selected for variants that were predicted to be deleterious by at least one of 6 prediction programs (PolyPhen-2 HDIV and HVAR, Mutation Taster, Mutation Assessor, FATHMM and RadialSVM), described in section 2.6.6.1. This meant that any differences between programs, would not lead to variants from being falsely discarded (Gray *et al.*, 2012). For example Tennessen *et al.* (2012) used four functional prediction programs and three conservation prediction methods, to see the overlap in the variant predictions. They found that using at least one tool, called approximately 74% of nonsynonymous variants deleterious and when using all 7, 1% of nonsynonymous variants were classed as functionally significant. It is possible that the method used here will keep numerous false positive variants, because they have been classed as pathogenic by only one program. In the future a more stringent method could be used, where a variant must be classed as deleterious by more than one prediction program. The exception to this step was that all INDELs were removed from the analysis. This was done because the accuracy of INDEL calling is very poor, where approximately 56.5% of INDEL calls are true (Fang *et al.*, 2014).

5.4.3. Variant frequency

This step aimed to further reduce the number of variants that were common in the general population. Since aHUS is a rare disease and it is hypothesised that the sequence variant causing disease, would also be rare. This step was performed using the MAF from the 1000g and ESP6500 databases. Variants with a MAF $\geq$1% in both population databases were removed from the analysis.

5.4.4. In silico conservation predictions

The final filtering step used was to positively select for variants that were found in highly evolutionary conserved sequence regions. This was determined using GERP++ and PhyloP. Variants with a GERP++ score of >2 and a PhyloP score >0.5 were positively selected. These thresholds were more conservative than those used in other studies (Dong *et al.*, 2015), yet sufficient to remove non-conserved variants defined as <0 for GERP++ (Davydov *et al.*, 2010) and <0 for PhyloP (Pollard *et al.*, 2010).

5.5. Candidate variant lists

The remaining 19 families with no known genetic cause of disease were run through the candidate selection pipeline. The analysis for family 8 was performed twice, first assuming just the mother was affected and secondly assuming both mother and daughter were affected. This was because the renal phenotype of the daughter had not been confirmed by biopsy and although she had proteinuria, she was noted to have diabetes mellitus, which is a known cause of proteinuria.

The results for dominant analysis are shown in Table 20. It can be seen that the number of variants in the final candidate list was far greater in some of the samples that have been WGA. For example families 13 and 14, have 7143 and 4147 variants respectively. This made analysis of these families, very difficult due to the large number of candidate variants. The average number of variants per family (excluding families 13 and 14) was 93, with a range of 1 to 335 variants. In some families there was only DNA available for testing in one individual. This made the analysis more difficult, because the final number of variants after filtering was higher, on average 164.4 variants compared to 66.1 variants (excluding WGA samples). In total 1589 variants were in the candidate variant list, consisting of a total of 1348 genes (excluding the data from families 13 and 14).

Family		Number of variants			
		Segregates	Deleterious	MAF <1%	Conserved
1		533	123	80	65
2		1178	291	203	170#
3		453	126	74	61
5*ǂ		5	1	1	1
6		1178	259	150	114#
8	II:2 affected	524	135	91	77#
	II:2 and III:1 affected	526	116	71	61
9		694	161	113	92
10		154	28	19	16
11*ǂ		319	76	51	39
12		625	141	76	58
13*ǂ		24622	7980	7940	7143#
14*		14377	4720	4634	4147#
15ǂ		513	148	110	95
16		571	114	78	58
17		1153	256	160	126#
19		455	112	68	60
20		686	145	102	87
21		597	137	91	74
25		1717	456	382	335#

Table 20 Variant table for a dominant mode of inheritance.

*Analysis for sequence variants that segregate following a dominant mode of inheritance. #= one family member tested using WES *= families containing WGA samples, ǂ=consanguineous families.*

Table 21 shows the combined list of variants from the autosomal recessive and X-linked modes of inheritance. There were fewer variants in the final candidate variant list compared to autosomal dominant inheritance analysis. 4 families were not run in this analysis, as they had multiple affected females across two generations, suggesting that it was not the result of an autosomal recessive or X-linked disorder. In total 43 genes had variants inherited in either an autosomal recessive or X-linked fashion, none of which segregated with more than one family.

Family		Number of variants			
		Segregates	Deleterious	MAF <1%	Conserved
1		8	0	0	0
2		52	13	11	4
3		8	5	4	2
5*ǂ		12	3	3	2
6		50	3	0	0
8	II:2 affected	16	1	1	1
	II:2 and III:1 affected	NA	NA	NA	NA
9		16	2	1	0
10		NA	NA	NA	NA
11*ǂ		11	3	2	2
12		20	0	0	0
13*ǂ		209	18	18	12
14*		50	11	7	6
15ǂ		40	18	15	10
16		NA	NA	NA	NA
17		25	7	4	2
19		32	2	0	0
20		15	3	3	2
21		15	1	1	0
25		NA	NA	NA	NA

Table 21 Variant table for a recessive mode of inheritance.

Analysis for sequence variants that segregate following an autosomal recessive or X-linked mode of inheritance. = families containing WGA samples, ǂ= consanguineous families, NA= Not available, indicates the families containing only affected females across several generations, which contraindicated this mode of inheritance.*

Compound heterozygous analysis was then performed, the results of which are shown in Table 22. For sequence variants to segregate with disease there must be more than one occurring within a gene, which was shared by all affected individuals. However it may be that only one of these variants would meet all the remaining criteria (deleterious, with a MAF of <1% and evolutionary conserved). 4 families were not analysed as they contained affected family members in multiple generations, which meant that according to our criteria, a compound heterozygous inheritance pattern was contraindicated. Families 13 and 14 contain WGA samples and found to have a far greater number of candidate variants compared to other families, on average 409.5 compared to 15.7 variants. Family 5 was also contained a WGA sample, however due to the large number of family members tested, no variants segregated with disease. In total there were 204 candidate variants found in 132 genes (excluding the data from families 13 and 14).

Family		Number of variants			
		Segregates	Deleterious	MAF <1%	Conserved
1		105	27	19	10
2		270	76	50	34
3		120	39	27	22
5*ǂ		0	0	0	0
6		297	91	61	35
8	II:2 affected	158	40	27	22
	II:2 and III:1 affected	NA	NA	NA	NA
9		155	35	20	12
10		19	1	0	0
11*ǂ		NA	NA	NA	NA
12		178	35	19	7
13*ǂ		612	406	393	328
14*		1075	608	582	491
15ǂ		28	16	11	8
16		NA	NA	NA	NA
17		248	60	38	24
19		62	18	10	7
20		165	38	25	12
21		99	27	19	11
25		NA	NA	NA	NA

Table 22 Variant table for a compound heterozygous mode of inheritance.

Analysis for sequence variants that segregate following a compound heterozygous mode of inheritance. indicates families containing WGA samples, ǂ indicates consanguineous families. NA= Not available, indicates the families containing affected family members in different generations, which contraindicated this mode of inheritance.*

The genes that contained candidate variants were then pooled for each mode of inheritance. The aim of which was to observe the number genes that are mutated in the different modes of inheritance. It would also enable the identification of genes that are mutated in more than one family, which may indicate a possible candidate for disease. The autosomal dominant and compound heterozygous analyses for families 13 and 14 generated a very large number of variants. This was predicted to be the result of WGA, which may have produced more sequencing errors. Therefore they were excluded from the analysis, to prevent dilution of meaningful candidates.

Figure 5-2 shows the results of the pooled gene lists for each inheritance analysis. Recessive inheritance (X-linked and autosomal recessive combined) saw the smallest number of genes in the remaining candidate list with 43 genes in total, all of which segregated in only one family. Compound heterozygous analysis identified 132 genes as

potential candidates in the 17 families investigated; 86% occurred in only 1 family, 9.8% in 2 families, 2.3% in 3 families and 1.5% in 4 families. Finally the dominant analysis identified 1202 genes; 89% occurred in only 1 family, 8.6% in 2 families, 1.6% in 3 families, 0.4% in 4 families and 0.15% in 5 families. In total 1359 genes segregated with 1 family, 129 genes with 2 families, 25 genes in 3 families and 8 genes in 4 families. Since aHUS is a very rare disease and the number of families available for testing was small, the large number of genes segregating in one family may not be surprising. If the exome data obtained in this project was combined with that of other international and European aHUS cohorts, it may identify genes that segregate with multiple families, which would not be seen in this study.

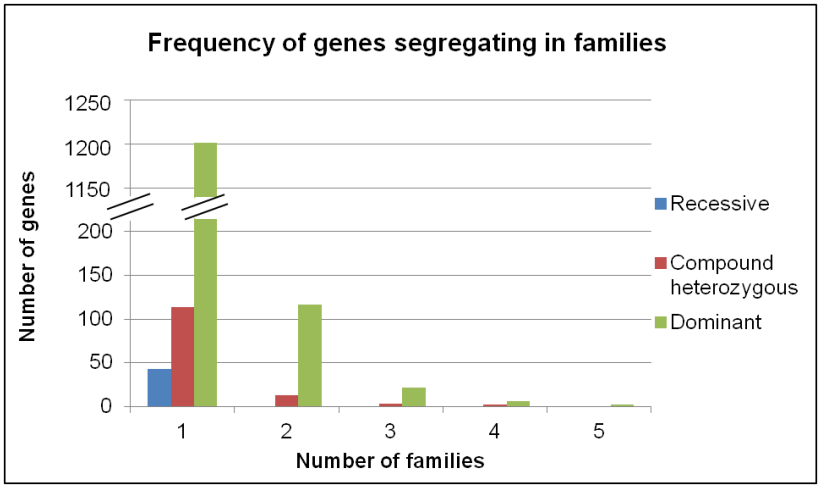

Figure 5-2 Frequency of candidate genes found in families.

Genes that were found by the candidate selection pipeline were pooled from all 19 families, to see if genes segregated in more than one family. The gene candidates for families 13 and 14 using the compound heterozygous and dominant analyses were excluded. The Y axis was broken, so the total number of genes identified in the dominant analysis in 1 family, was 1202.

Two genes were found in 5 families, which were *GPR98* and *RGPD3*. The variant p.T4090N, in *GPR98*, was seen in 4 out of 5 families. It seemed very unlikely that a single variant would be seen so frequently in a rare disease, leading to the hypothesis that this may be a sequencing artefact. Moreover, review of the literature indicated that mutations in *GPR98* were associated with Usher syndrome, an autosomal recessive disorder, characterised by deafness and blindness (Ebermann *et al.*, 2009). Due to the absence of those phenotypes in the families described here, *GPR98* was not considered for further analysis. *RGPD3* had minimal information available that could suggest a possible mechanism of disease; therefore this gene was not prioritised for further

analysis in this project. However due to their fulfilment of the search criteria, they could be considered for further analysis in the future.

The gene lists for recessive (autosomal recessive and X-linked) and compound heterozygous analyses were then pooled, in order to determine if these analyses had any common genes, suggesting that a functional deficiency would cause disease. In total 132 genes had variants that occurred in compound heterozygous analysis and 43 genes had variants that were observed in an autosomal recessive or X-linked analysis. Only 1 gene was found that was found in both analyses, shown in Figure 5-3. This gene was Diacylglycerol kinase ε (*DGKE),* described in section 5.6.

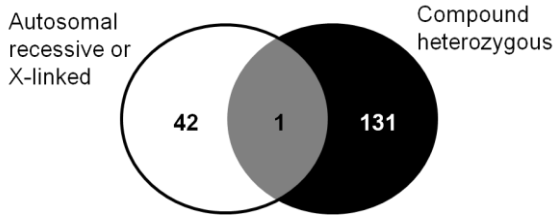

Figure 5-3 Venn diagram of genes identified in recessive and compound heterozygous analyses.

Genes containing variants that were inherited in the recessive analysis (autosomal recessive and X-linked combined) were shown in white. Genes with compound heterozygous variants were shown in black. Genes containing variants found in both analyses are in the overlapping grey section.

5.6. Diacylglycerol kinase epsilon (DGKE)

5.6.1. Gene

Diacylglycerol kinases (DGKs) are important in regulating the formation of Diacylglycerol (DAG), a component of an intracellular lipid signalling pathway known as the Phosphatidylinositol (PI) cycle (Shulga *et al.*, 2011a). The PI cycle is based on the hydrolysis of phosphatidylinositol (4,5)-bisphosphate (PIP_2), by Phospholipase C (PLC) into DAG (Dolley *et al.*, 2009). DAG is then able to activate PKC (Protein Kinase C), leading to many downstream effects including changes in vascular tone (Dolley *et al.*, 2009, Khalil, 2013), the release of prothrombotic factors (Kikkawa *et al.*, 1989, Lemaire *et al.*, 2013) and changes to the actin cytoskeleton (Brandt *et al.*, 2002). The role of DGKs is to regulate the formation of DAG. They can do this by phosphorylating DAG, converting it into phosphatidic acid (PA).

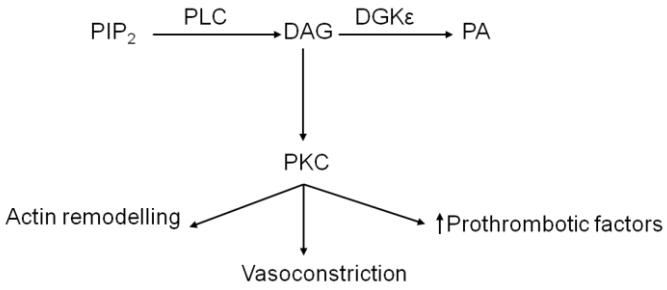

Figure 5-4 Schematic diagram of PI pathway.

PLC converts PIP_2 into DAG, which can then activate PKC, leading to downstream effects such as actin remodelling, vasoconstriction and increased expression of prothrombotic factors. DGKε prevents this by converting DAG into PA. Adapted from Lemaire et al. (2013).

There are currently ten described DGKs, divided into 5 subtypes that share a common protein structure (Shulga *et al.*, 2011a). Diacylglycerol kinase epsilon (DGKε) is a type III kinase, encoded by *DGKE* located on chromosome 17q22 (Hart *et al.*, 1999). It is a 64kDa protein composed of 5 domains, including an N-terminal hydrophobic domain (HD), two cysteine-rich (Cys1-1 or 2) domains, a kinase catalytic domain (KCD) and a C-terminal kinase accessory domain (KAD) shown in Figure 5-5. It is the only DGK that lacks a regulatory motif, suggesting that DGKε is constitutively active (Topham and Epand, 2009). DGKε is localised to the endoplasmic reticulum (ER) and the plasma membrane (Decaffmeyer *et al.*, 2008). It has been shown to be expressed in podocytes (Ozaltin *et al.*, 2012), platelets (Yada *et al.*, 1990) and glomerular endothelial cells (Lemaire *et al.*, 2013).

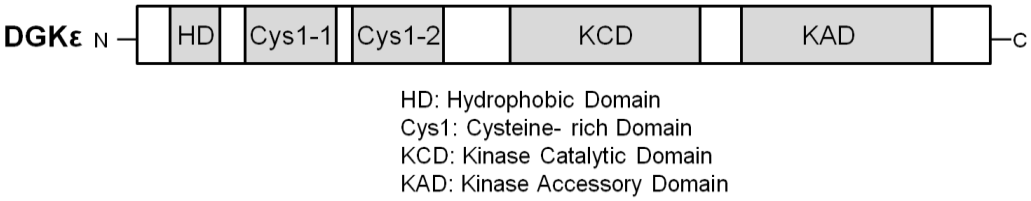

Figure 5-5 Protein structure of DGKε.

Adapted from Lemaire et al. (2013)

The HD domain contains approximately 20 hydrophobic residues that allow DGKε to interact with the intracellular aspect of the cell membrane (Decaffmeyer *et al.*, 2008) and the ER (Matsui *et al.*, 2014). The precise function of Cys1 domains is not currently known. However they contain zinc finger-like motifs and are thought to be involved in DAG-binding (Shulga *et al.*, 2011a). KCD has a region that mediates ATP binding, which is critical for DGK function (Sakane *et al.*, 1996). Studies have suggested that the

KCD and KAD domains are not able to function alone, also requiring the other domains to provide functional kinase activity (Rodriguez de Turco *et al.*, 2001).

Unlike other DGKs, DGKε has specificity to certain DAGs containing arachidonoyl chains at the *sn*-2 position (Shulga *et al.*, 2011b), also known as 1-stearoyl-2-arachidonoyl glycerol (Tang *et al.*, 1996).

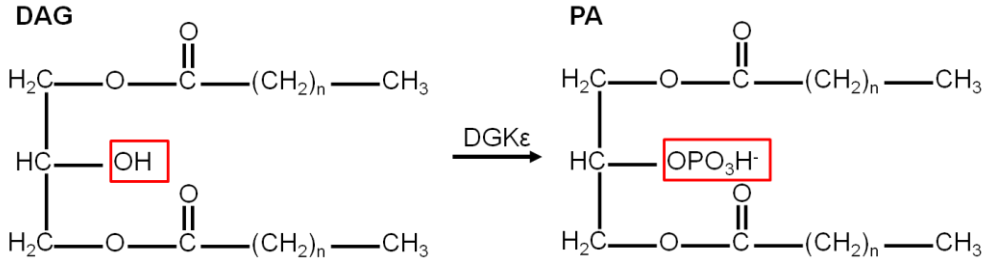

Figure 5-6 Phosphorylation of DAG into PA by DGKε.

DGKε specifically targets DAG containing an arachidonoyl chain at the sn-2 position. When it converts DAG into PA the OH group highlighted in the red box is phosphorylated. Adapted from Shulga et al. (2011a).

This specificity was hypothesised to be due to the presence of an amino acid motif at position 431-443, $LX_{(3-4)}RX_{(2)}LX_{(4)}G$ found within the KAD (Shulga *et al.*, 2011b). It was thought that amino acids L431 and L438 form the bottom of the DAG-binding pocket and their mutation reduced binding affinity (Shulga *et al.*, 2011b). This section described the sequence variants found in *DGKE* in this project.

5.6.2. Family 20 and 8

Family 20 has two affected siblings with recurrent aHUS. The affected proband had a biopsy at 4 years of age confirming TMA. There were also features of membranoproliferative glomerulonephritis (MPGN) with C3 deposition (+++), IgM (++) with little IgG on immunofluorescence. His brother (II:2) was biopsied at 13 years of age that demonstrated a MPGN with no C3 deposition, IgM (++) with little IgG on immunofluorescence. There was also evidence of secondary focal segmental glomerulosclerosis. Complement levels were measured at 27 years of age and C3 (1.25g/L), C4 (0.33g/L), FH (0.57g/L) and FI (51mg/L) and found to be within the normal range and MCP expression was normal.

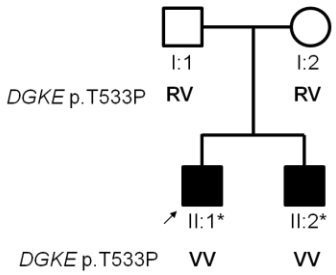

Figure 5-7 Pedigree of family 20.

The proband is indicated by an arrow. '' indicated samples tested by WES. Genotype was indicated by RV (heterozygous) or VV (homozygous).*

Family 8 has two affected individuals in the second generation, one of whom is now deceased (II:3). The affected proband (II:2) had aHUS at the age of two, which was confirmed by a renal biopsy. Dialysis was not required. Complement levels in serum were normal (C3 1.01g/L, FH 0.55g/L, FI 57mg/L), except for C4, which was low 0.14g/L (normal 0.18-0.6g/L). Routine Sanger sequencing of *MCP*, *CFH* and *CFI* did not reveal any disease causing variants. The proband's sister (II:3) died of aHUS at the age of three, no DNA was available for testing. The proband's offspring (III:1) did not have a renal biopsy, but had proteinuria which was thought to be due to poorly controlled diabetes mellitus. Her kidney function was said to be normal. Due to the lack of definitive evidence, analysis was undertaken assuming both affected and unaffected status

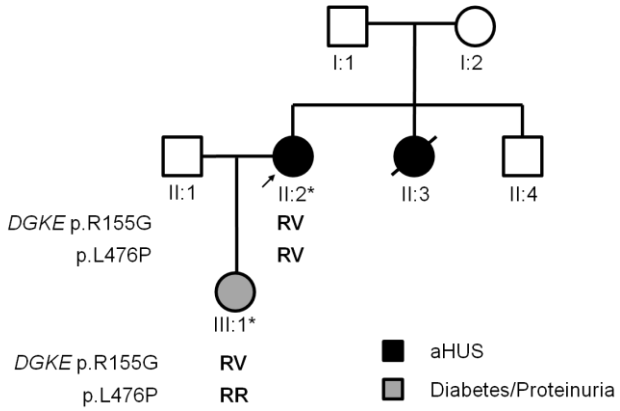

Figure 5-8 Pedigree of family 8.

Proband is indicated by the arrow and '' indicates all family members tested by WES. The proband and deceased sibling (II:3) had confirmed aHUS, shown by black circles. The daughter III:1 has diabetes and proteinuria, shown in grey. Genotype was shown in bold.*

WES analysis was carried out using several modes of inheritance (autosomal recessive and X-linked, autosomal dominant and compound heterozygous) and candidate variants

were pooled to see if any gene had variants in multiple families. Using this method, *DGKE* was found in the autosomal recessive/X-linked and compound heterozygous analyses, suggesting that complete functional deficiency led to disease. Family 20 had a homozygous change c.1597A>C, p.T533P, located in exon 12, which was present in both affected offspring, II:1 and II:2. Both parents were confirmed to be heterozygous. Family 8 had a compound heterozygous change, found only in the proband (II:2). The changes found were c.1427T>C, p.L476P, occurring in exon 11 and c.463A>G, p.R155G in exon 2. Only R155G was seen in the daughter III:1, suggesting that the two mutations were found on separate alleles. Due to her lack of compound heterozygosity, III:1 was now treated as unaffected. Table 23 shows that all three variants reported in *DGKE* were not seen in the 1000g or ESP6500 databases.

Family	Sequence variant		Exon	Genotype	MAF %
	Transcript	Protein			
20	c.1597A>C	p.T533P	12	VV	0
8	c.463A>G	p.R155G	2	RV	0
	c.1427T>C	p.L476P	11	RV	0

Table 23 DGKE variant data in families 20 and 8.

DGKE variant data in Families 20 and the affected proband in family 8 (II:2). RV= indicates that patients had one copy of the reference and mutant alleles. VV= indicates that patients were homozygous for the mutant allele. MAF frequency listed is from 1000g project and ESP6500.

These variants were then confirmed by Sanger sequencing. Figure 5-9 shows that the parents of family 20, were heterozygous for T533P and the affected proband (II:1) was homozygous, the trace for II:2 was not shown.

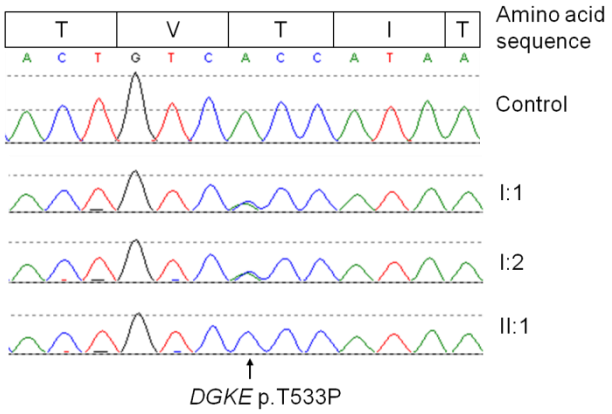

Figure 5-9 Sanger sequencing trace for family 20.

Top trace is the control, showing a homozygous c.1597A, p.T533. The next two traces are for the mother and father, showing a heterozygous c.1597A>C, causing an amino acid substitution T533P. Finally the trace for the affected proband, II:2, shows homozygosity for T533P. The amino acid sequence encoded by exon 12 is shown above.

Figure 5-10 shows the Sanger trace for the proband in family 8. The results of the unaffected offspring (III:1) were not shown.

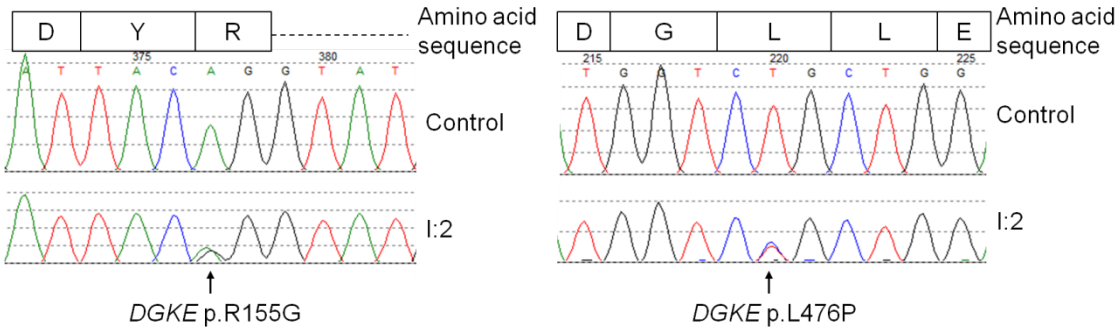

Figure 5-10 Sanger sequencing trace for family 8.

Left chromatogram shows sequencing trace of DGKE R155G and the right for DGKE L476P. Top trace is the control and the bottom trace is for the proband of family 8 (I:2). The amino acid sequence encoded by the exon is shown above, where the dotted line denotes the intronic region.

The finding of two families with *DGKE* variants, led to the review of all exome data for any other *DGKE* sequence variants. This led to the finding that there was a heterozygous change in *DGKE*, in Family 13. This was c.826delG, p.V276FfsX8, which was not seen in 1000g, ESP6500 or dbSNP databases. This variant was removed during variant filtering, because it was an INDEL.

This was a consanguineous family, where the parents are first cousins (Kaplan *et al.*, 1992), shown in Figure 5-11. There have been 4 affected offspring, 3 of whom are now

deceased. There was no history of aHUS in previous generations. The proband was treated with plasma exchange and Sanger sequencing of *CFH, MCP* and *CFI* revealed no sequence variants.

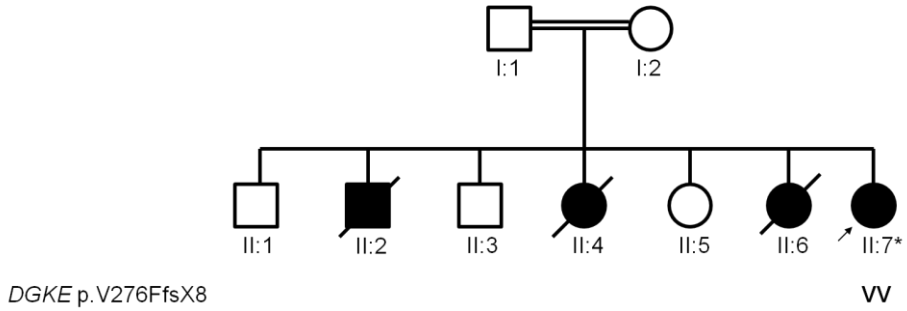

Figure 5-11 Pedigree of family 13.

Proband is indicated by the arrow and '' indicates all members tested by WES. The double band between parents I:1 and I:2, indicates that there is consanguinity. Genotype was shown in bold.*

With the knowledge that this family is consanguineous and that the DNA sample for the proband was WGA and therefore poor quality, it was hypothesized that this variant may have been poorly covered by the exome sequencing and that the proband may in fact be homozygous. Therefore Sanger sequencing was carried out on non-amplified DNA, which led to the confirmation that the patient is homozygous for V276FfsX8, shown in Figure 5-12.

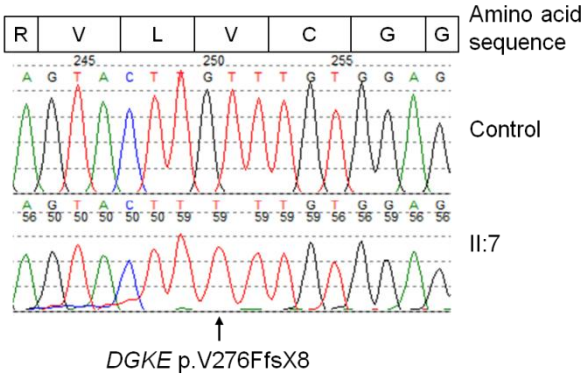

Figure 5-12 Sanger sequencing trace for family 13.

The top trace shows the result of the control, the bottom trace shows the result for the affected proband, II:7, shows homozygosity for V276FfsX8. The amino acid sequence encoded by the exon is shown above.

At this point there were two publications that supported the findings described here, which was that complete functional deficiency of DGKε was associated with aHUS

(Lemaire *et al.*, 2013, Ozaltin *et al.*, 2012). There was then report of an intronic variant in *DGKE*, c.888+40A>G, which was associated with disease and reported to cause aberrant splicing of exon 5 (Mele *et al.*, 2015). The WES coverage for this position was poor therefore Sanger sequencing primers were designed to analyse this region. All families with an unknown genetic cause of disease were analysed. However this variant was not identified in any patients.

5.6.3. Newcastle aHUS cohort rescreening

It was observed in other cohorts and also in the data presented here, that all affected individuals with DGKε deficiency, presented with disease early in life and had disease recurrence during eculizumab treatment (Ozaltin *et al.*, 2012, Lemaire *et al.*, 2013, Sanchez Chinchilla *et al.*, 2014). Therefore it was hypothesised that there may be additional patients within the Newcastle cohort with DGKε deficiency. These patients would not respond to eculizumab treatment (Lemaire *et al.*, 2013) and therefore it may be prudent to withdraw treatment, due to its high cost and its associated risk to meningococcal infections. All sporadic paediatric and adults cases, who were being treated with eculizumab were screened for all exons of *DGKE*, including the intronic variant c.888+40A>G. In total 52 paediatric and 61 adults cases were screened, which identified three sporadic paediatric cases with *DGKE* sequence variants, all of which were present in homozygosity. Sporadic patients 1-3 were referred to as Sp1-3.

Sp1 was born of consanguineous parents. Renal biopsy of the proband confirmed TMA and also demonstrated features of MPGN. Complement levels were normal (C3 1.01g/L and C4 0.22g/L).

Sp2 presented with HUS at 7 months, which occurred after a diarrhoea illness. There was no evidence of *E. coli* O157 in her stools or serology. She was treated with peritoneal dialysis for two weeks, during which time she had oligoanuria and thrombocytopenia. This was followed by two further disease relapses within one year, during which time she was treated with fresh frozen plasma. At 8 years old she had chronic kidney disease, with a marginally elevated creatinine of 69µmol/L, proteinuria and hypertension. ADAMTS13 activity was 33%, excluding the diagnosis of TTP. Complement levels (C3 1.64g/L, C4 0.81g/L, FH 0.72g/L and FI 69mg/L) and MCP expression were normal.

Sp3 is a child of Middle East origin, who presented with aHUS at 8 months of age and was treated with plasma therapy, whilst on this treatment the patient had multiple disease relapses. Eculizumab treatment was initiated at 6 years of age and administered for 5 years. During this time the patient relapsed on two occasions. At age 11 she had chronic kidney disease and continued to have nephrotic syndrome whilst on eculizumab treatment.

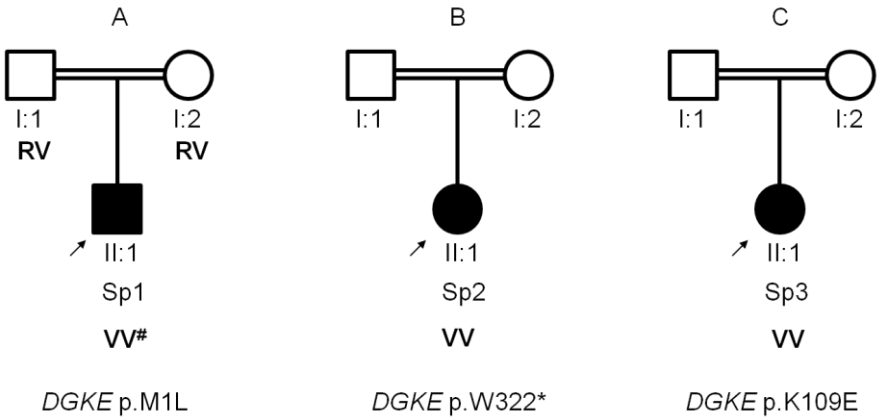

Figure 5-13 Pedigrees for Sporadic aHUS patients with DGKE mutations.

A-Pedigree for Sp1, B-Pedigree for Sp2 and C-Pedigree for Sp3. The proband is indicated by an arrow. '#' indicates predicted genotype.

Sequence variants were confirmed by Sanger sequencing in Sp2 and Sp3, shown in Figure 5-14B and C respectively. There was no DNA available for Sp1, so Sanger sequencing was performed on the parents, shown in Figure 5-14A. The parents both carried the mutant allele in heterozygosity, which led to the hypothesis that the affected offspring had inherited the allele from each parent and therefore was homozygous for M1L.

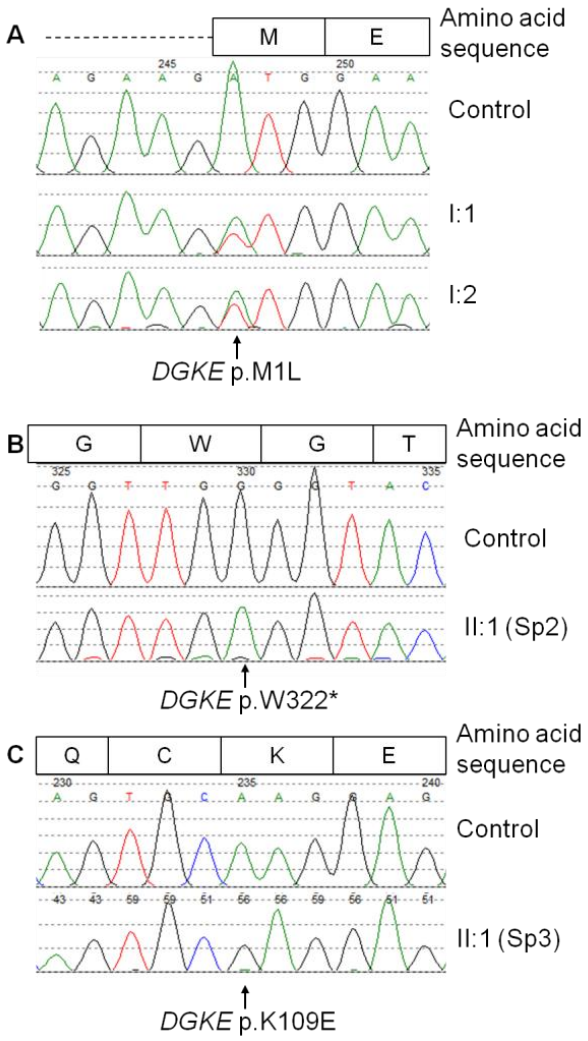

Figure 5-14 Sanger traces for Sporadic patients.

A- The parents of the proband, Sp1. B- Sp2. C- Sp3. The amino acid sequence encoded by the exon is shown above the sequence chromatogram, where the dotted line denotes the intronic region.

M1L and K109E were not present in either 1000g or ESP6500 databases. W322* is listed on dbSNP (rs138924661) and had a MAF in ESP6500 of 0.015%; however it was not found in 1000g database. W322* was reported by Lemaire *et al.* (2013) and Sanchez Chinchilla *et al.* (2014), both in homozygosity and compound heterozygosity, in aHUS patients.

Patient ID	Sequence variant		Exon	Genotype	MAF %
	Transcript	Protein			
Sp1	c.1A>T	p.M1L	1	VV‡	0
Sp2	c.966G>A	p.W322*	6	VV	0.015
Sp3	c.325A>G	p.K109E	2	VV	0

Table 24 DGKE variants observed in sporadic paediatric cases.

VV= indicates that patients were homozygous for the mutant allele. '‡' Indicates the predicted genotype as no DNA was available for testing. MAF frequency listed is from ESP6500 as no variants were reported in 1000g.

All *DGKE* variants reported in this project were nearly entirely absent from the normal population. The only exception is W322*, which was seen in ESP6500. However this was still very rare (MAF of 0.015%) and the occurrence of a premature stop codon is highly likely to be pathogenic. This suggested that these sequence variants were likely to be functionally significant. Consequently functional analysis was performed to see if they were predicted to be pathogenic.

5.6.4. In silico analysis

To examine the effects of these sequence variants on the protein, several *in silico* tools were used. V276FfsX8 was not analysed because none of the programs used were able to formulate predictions for INDELs. However the formation of a premature stop codon was considered to be severely detrimental.

The amino acid substitution M1L resulted in the loss of the initiating methionine and as a result, it was predicted to disrupt protein translation. Even so M1L was only predicted to be deleterious in 1 out of 4 tools, showing the limitations of these tools. The remaining *DGKE* variants were predicted to be deleterious by at least one *in silico* tool, shown in Table 25.

Patient ID	Variant	*In silico* testing					
		PolyPhen-2		Mutation Taster	Mutation Assessor	FATHMM	Radial SVM
		HDIV	HVAR				
Family 20	T533P	D	P	D	L	T	T
Family 8	R155G	B	P	D	M	D	D
	L476P	D	D	D	M	T	T
Sp1	M1L	B	B	D	NA	T	NA
Sp2	W322*	NA	NA	D	NA	NA	NA
Sp3	K109E	D	D	D	M	D	NA

Table 25 In silico data for DGKE variants

Table shows the in silico predictions for the DGKE variants identified. Software not able to calculate predictions were labelled as NA, not available. B= Benign, D= deleterious, L= Low, M= medium, P= polymorphism, T= Tolerated. Scores could not be calculated for V276FfsX8.

One other mechanism to determine if a variant might be functionally significant, was to see if it occurred in a region of DNA that was evolutionary conserved. If it was maintained across several species then it indicated an importance in function. To examine if these variants occur in conserved areas, the human protein sequence of DGKε, was compared to 10 other species. It can be seen in Figure 5-15 that all 5 variants found in the Newcastle cohort were found to occur in highly conserved regions. With the exception of M1, all amino acids were maintained through to zebrafish.

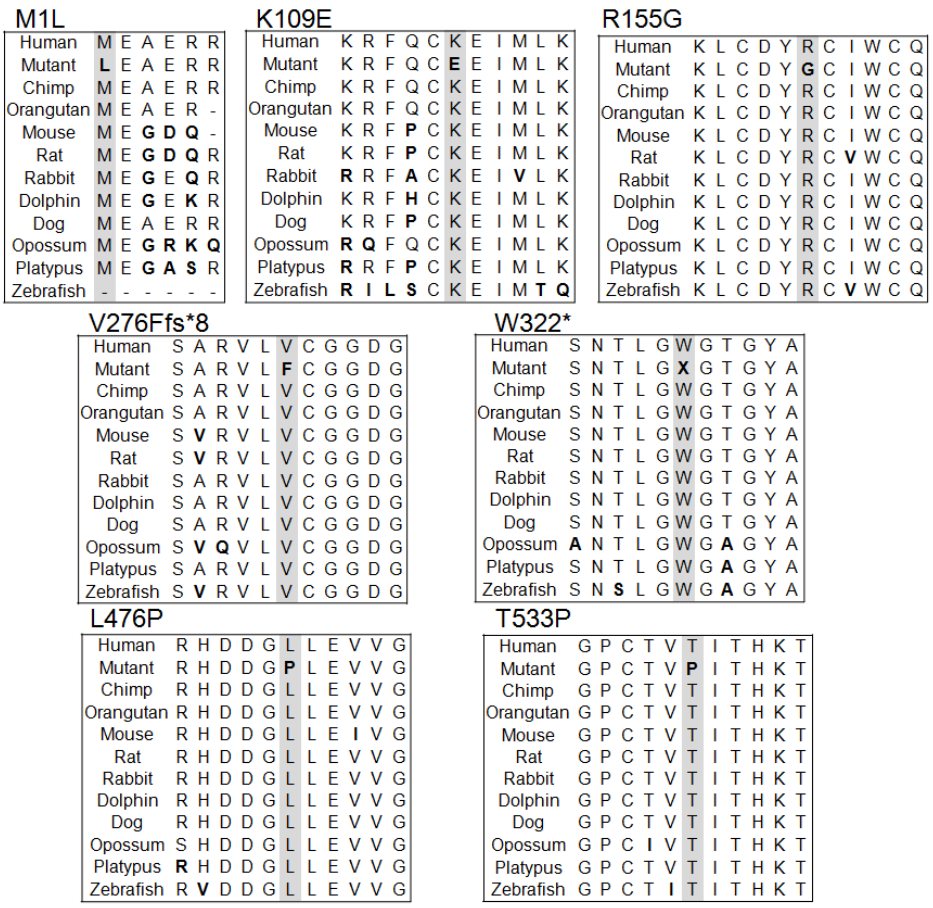

Figure 5-15 Amino acid sequence alignment for DGKE variants.

Amino acid sequence alignment for sequence variants in DGKE. Human with and without the variant are compared to 10 other species. Highlight grey is the position of the variant described here. Amino acids in bold, indicate differences to the reference sequence.

These results were expected as the conservation scores from GERP++ and PhyloP were above our preset threshold of 2 and 0.5 respectively. Table 26 shows the values obtained using GERP++ and PhyloP. Scores for M1L, W322* and K109E could not be calculated using GERP++, as this was not a web-based platform. The score for V276FfsX8 was unavailable as neither GERP++ nor PhyloP could calculate scores for INDELs.

Patient ID	Variant	GERP++	PhyloP
Family 20	T533P	5.80	2.216
Family 8	R155G	4.56	0.936
	L476P	5.36	2.15
Sp1	M1L	NA	3.313
Sp2	W322*	NA	5.709
Sp3	K109E	NA	4.553

Table 26 Conservation scores for DGKE variants.

NA= Not available. Scores could not be calculated for V276FfsX8.

Analysis of variants using a combination of *in silico* tools demonstrated that all variants found in *DGKE*, in Newcastle patients, were predicted to be deleterious by at least 1 prediction tool and disrupted highly conserved amino acids. V276FfsX8 could not be analysed with prediction software used. However due to the occurrence of a premature stop codon, it was predicted to be pathogenic. In addition amino acid sequence alignment demonstrated that this was an evolutionary conserved area, indicating it may have a critical function.

5.6.5. Protein modelling

To investigate the effect of the point mutations K109E, R155G, L476P and T533P on DGKε, protein modelling was carried out. A predicted protein structure of DGKε was obtained using Phyre2 and manipulated in PyMOL. K109E was positioned close to the HD, Cys1-1 and KCD domains, suggesting that it might have a subtle effect on the function of all three domains, which might include disrupting DGKε localisation to the cell membrane and the ER, altering affinity to DAG or altering ATP binding. R155G was positioned on the surface of the Cys1-2 domain, the function of which is not currently known. L476P and T533P were in close proximity to the two amino acids, which have been previously demonstrated to be required for DAG arachidonoyl specificity (Shulga *et al.*, 2011b). However it is possible that these variants cause disease by preventing the expression of DGKε.

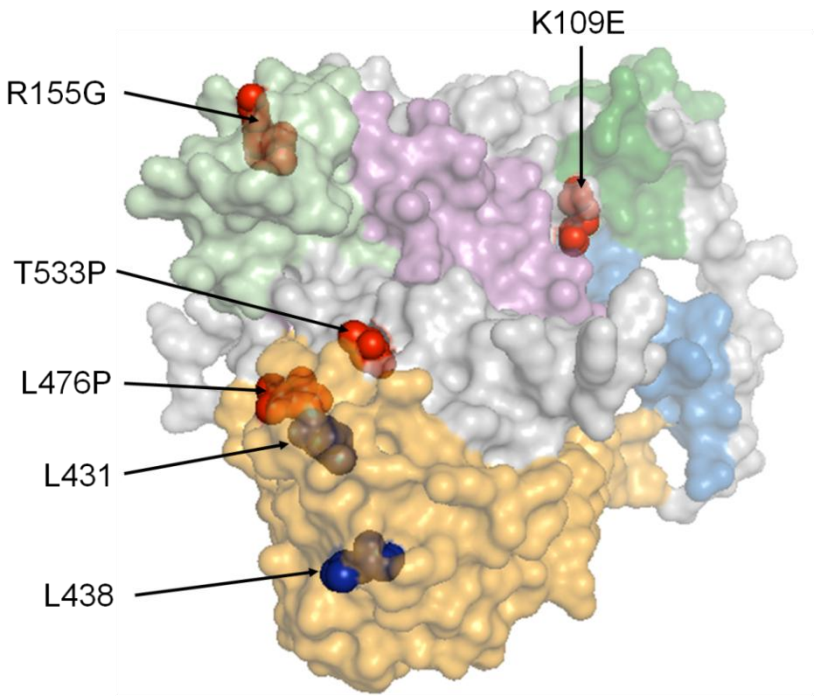

Figure 5-16 Predicted DGKε protein model.

Phyre2 was used to create a predicted protein model of DGKε. Protein shown in surface mode, at a transparency setting of 0.4. Amino acids L431 and L438, thought to important for DAG specificity are shown blue spheres. Amino acid substitutions identified in this project are shown by red spheres. HD (amino acids 22-42) shown in blue, Cys1-1 (amino acids 60-109) shown in dark green, Cys1-2 (amino acids 125-178) shown in light green, KCD (amino acids 219-350) shown in pink and KAD (amino acids 369-524) shown in orange.

5.6.6. Pathophysiology

At the time this project was started there were no reported genes associated with aHUS that were outside the complement system, with the exception of *THBD* and *MMACHC*. However there is now increasing evidence from other European cohorts that DGKε deficiency is associated with early onset aHUS. Figure 5-17 demonstrates the sequence variants in DGKε, reported by other cohorts and those described here. It can be seen that sequence variants occurred across the entire length of the protein and all variants reported here were novel, apart from W322*, which was also reported by Lemaire *et al.* (2013) and Sanchez Chinchilla *et al.* (2014).

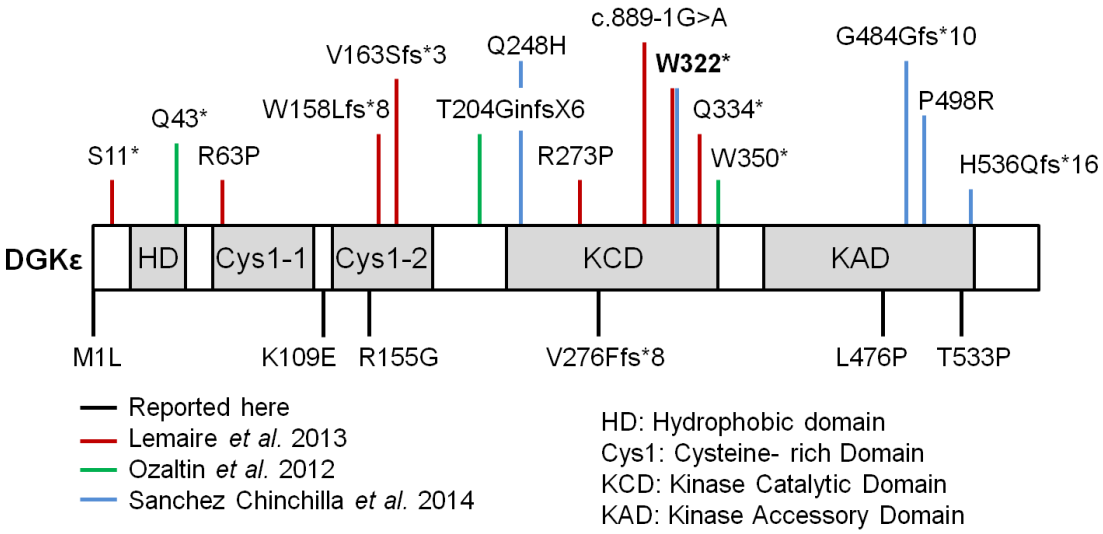

Figure 5-17 Diagram showing mutations reported in DGKE.

Green lines are from Ozaltin et al. (2012). Red lines are from Lemaire et al. (2013). Blue lines are from Sanchez Chinchilla et al. (2014). Black lines are reported here. W322 is in bold, indicating it was found in the Newcastle cohort, but it had been perviously reported. All variants described here were homozygous, with the exception of R155G and L476P, which were found in compound heterozygosity. Added in proof: another paediatric case has been demonstrated to be compound heterozygous for DGKE mutations (c.465-2A>G and c.393C>G, p.N131K).*

The current mechanism of disease is unknown, but evidence had suggested that the loss of DGKε was distinct from the complement system (Bruneau *et al.*, 2014). This was not surprising as patients with an absence of functional DGKε, did not respond to eculizumab, a complement-directed therapy (Rother *et al.*, 2007). It was proposed that the accumulation of DAG would lead to increased PKC signalling, the downstream effects of which would lead to aHUS (Noris *et al.*, 2015), shown in Figure 5-18.

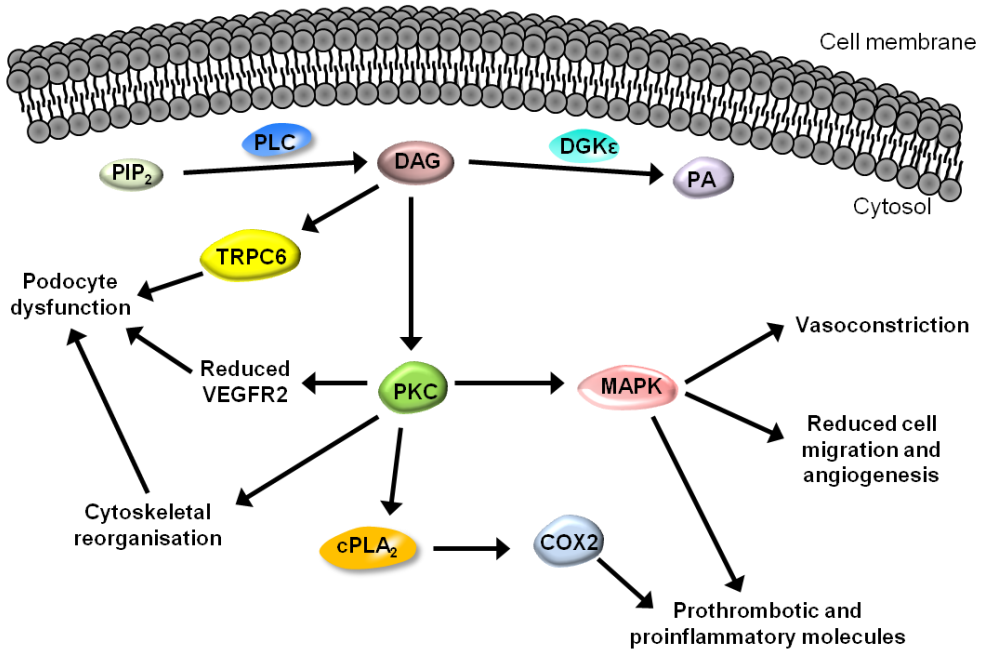

Figure 5-18 DAG signalling pathway.

PLC converts PIP_2 into DAG, which has many downstream signalling targets, such as TRPC6 and PKC. DGKε regulates this pathway by converting DAG into PA. Adapted from Noris et al. (2015).

One downstream effect of PKC signalling was the activation of p38- Mitogen-activated protein kinase (MAPK) (Vijayan, 2015, Bruneau *et al.*, 2014). Bruneau *et al.* (2014) demonstrated using HUVECs and human dermal microvascular endothelial cells (HMECs) that p38-MAPK signalling created a prothrombotic environment. This was due to the up regulation of Intercellular Adhesion Molecule 1 (ICAM-1), E-selectin, tissue factor and the reduced migration and angiogenic responses of endothelial cells. PKC-mediated activation of MAPK and MAPK kinase (MEK) was also demonstrated to play a role in vasoconstriction (Khalil, 2013).

PKC signalling also increased the production of prothrombotic mediators such as vWF (Carew *et al.*, 1992) and plasminogen activator inhibitor-1 (PAI-1) (Ren *et al.*, 2000). PKC can also activate cytosolic phospholipase A_2 ($cPLA_2$), leading to arachidonic acid production, which can be converted into prostaglandins by cyclooxygenases (COX) 1 or 2 (Noris *et al.*, 2015). PKC has been shown to down regulate VEGF receptor (VEGFR) 2 in porcine aortic endothelial cells, by targeting it for degradation (Singh *et al.*, 2005). VEGF inhibitors are known to cause TMA due to damage to glomerular endothelial cells, altering their fenestrated structure (Eremina *et al.*, 2008). Therefore DGKε

deficiency may lead to decreased VEGF signalling, causing vascular endothelial cell damage, similar to what is seen in VEGF inhibitor-mediated aHUS.

Studies have shown that elevated DAG can lead to increased activity of a cation channel Transient receptor potential cation channel, subfamily C, member 6 (TRPC6) (Ozaltin *et al.*, 2012, Hofmann *et al.*, 1999) TRPC6 is found in the podocyte slit diaphragm and activating mutations have been associated with FSGS (focal segmental glomerulosclerosis) (Winn *et al.*, 2005). This is the result of increased intracellular Ca^{2+} that causes podocyte dysfunction and thus disruption of the glomerular filtration barrier (GFB) (Heeringa *et al.*, 2009). Patients with *DGKE* mutations also get proteinuria between thrombotic events, which has been suggested to be caused by a similar mechanism (Noris *et al.*, 2015).

Down regulation of *DGKE* was shown by Matsui *et al.* (2014), to make cells more vulnerable to apoptosis as a result of ER stress. The use of DGK inhibitor R59022, led to increased platelet secretion and aggregation in response to thrombin (Nunn and Watson, 1987). This suggested that loss of functional DGKε could cause podocyte damage and create a prothrombotic environment in the glomerulus. Mouse models have shown that knocking out *DGKE*, slows the PI pathway. These mice are more tolerant to seizures caused by electroconvulsive shocks but do not develop a renal phenotype, although this may be the result of differential localisation of DGKε between mice and humans (Rodriguez de Turco *et al.*, 2001).

One study identified DGKε localisation to actin stress fibres in rat aortic smooth muscle, suggesting it plays a role in cytoskeletal regulation (Nakano *et al.*, 2009). Indeed PKC has been shown to be involved in cytoskeletal rearrangements, specifically the disassembly of the stress fibres in vascular smooth muscle cells (Brandt *et al.*, 2002). Thus one hypothesis could be that the loss of DGKε resulted in increased PKC-mediated stress fibre disassembly, potentially affecting the integrity of the GFB.

The study by Sanchez Chinchilla *et al.* (2014) identified the coinheritance of *THBD* and *C3* mutations in patients with *DGKE* sequence variants. The *THBD* mutation observed had been previously reported in association with aHUS (Delvaeye *et al.*, 2009) and the *C3* mutation described was novel and predicted to be deleterious. The patients with the *THBD* and *C3* mutations had a more severe phenotype with increased disease recurrence. It was also noted that the patient with a *C3* mutation had a positive response

to eculizumab therapy. These results suggested mutations in genes known to cause aHUS may increase disease severity. No mutations in complement genes or *THBD* were seen in concomitance with *DGKE* sequence variants.

M1L leads to loss of the initiating methionine that was hypothesised to lead to failure of protein translation. K109E was located between the two Cys1 domains and R155G occurred in the second Cys1 domain. The functions of the Cys1 domains are still unclear; however it is thought they are involved in DAG binding, therefore these variants may disrupt this interaction. L476P and T533P were both located in the KAD and therefore could disrupt DGKε's arachidonoyl chain-specificity, due to their close proximity to the DAG recognition motif (Shulga *et al.*, 2011b). The occurrence of premature stop codons in V276FfsX8 and W322* were predicted to lead to either truncated proteins or nonsense-mediated mRNA decay and thus DGKε deficiency. Consequently all these variants were hypothesized to lead to absence of functional DGKε activity. This would cause uncontrolled DAG signalling that was predicted to be prothrombotic, leading to disease onset. It would be important to try and test DGKε expression in biopsy tissue from affected individuals to test if whether or not these sequence variants abolish protein expression. If there was protein expression then it would be necessary to undertake functional analysis to determine their potential effects. Finally further work is needed to elucidate the precise mechanism in which *DGKE* mutations cause disease.

5.7. Inverted formin 2 (INF2)

Inverted formin 2 (INF2) is a regulator of actin polymerisation and depolymerisation (Chhabra and Higgs, 2006). It is encoded by *INF2* which is found on chromosome 14 and encoded by 22 exons (ENST00000330634). This section described sequence variants identified in *INF2* in familial aHUS cases.

5.7.1. Gene function

Actin filaments are linear, double helix structures that are composed of many actin monomers (Alberts *et al.*, 2002). The actin cytoskeleton is involved in many cellular processes such as cell migration, phagocytosis, immunological synapses and adherens junctions (Chhabra and Higgs, 2007). Filaments are constantly being formed, with new actin monomers being added and removed from the ends of filaments. Monomer

addition occurs at a greater rate at the barbed end and monomer removal occurs more at the pointed end, creating a 'treadmilling' effect (Alberts *et al.*, 2002). Once actin monomers have joined Adenosine triphosphate (ATP) is hydrolysed into Adenosine diphosphate (ADP), to form a stable filament (Lodish *et al.*, 2000), demonstrated in Figure 5-19. This process is regulated by associated proteins, such as the Arp2/3 complex, spire proteins and formins (Chhabra and Higgs, 2007).

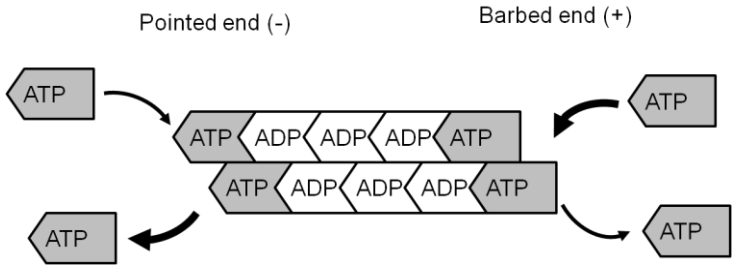

Figure 5-19 Schematic diagram of actin filament assembly.

ATP- actin monomers are incorporated or lost at the pointed or barbed ends of actin filaments. Once they are bound ATP is hydrolysed to form ADP. Grey shapes indicate ATP-actin monomers. White shapes indicate ADP-actin monomers. The larger arrows indicate the preferential addition or loss of actin subunits. Adapted from Lodish et al. (2000).

Formins are a group of proteins that are involved in the maintenance and regulation of actin and tubulin cytoskeletal networks (Goode and Eck, 2007). Formins produce actin filaments that are not branched, unlike the action of Arp2/3 complexes and they are also able to elongate actin filaments, unlike Spire proteins (Chhabra and Higgs, 2007). In mammals there are 15 formin genes, subdivided into 7 subgroups (Goode and Eck, 2007). They share a characteristic structure with an N-terminal GTPase-binding domain (GBD), followed by the Diaphanous Inhibitory Domain (DID) domain, Formin homology 1 and 2 (FH1 and FH2) domains and then a Diaphanous Autoregulatory Domain (DAD) domain at the C-terminus (Breitsprecher and Goode, 2013). They are able to regulate themselves by autoinhibition, due to the interaction of the DID and DAD domains (Breitsprecher and Faix, 2010). They have several functions such as actin nucleation (Gould *et al.*, 2011) and filament elongation (Chesarone *et al.*, 2010). Inverted formins (INFs) form one subgroup of formins and includes INF2.

INF2 has a similar structure to other formins although it lacks a GBD and has a WH2 (Wiskott-Aldrich syndrome protein-homology 2) motif within the DAD (Chhabra and Higgs, 2006). The FH1 domain binds to profilin, a protein which sequesters actin monomers, accelerating the polymerisation activity of the FH2 domain (Kovar *et al.*,

2006). The FH2 domains dimerise and encircle the filaments, accelerating nucleation and elongation of the barbed ends of filaments (Gurel *et al.*, 2014). The C-terminal WH2 domain has several functions. Firstly in conjunction with the FH2 domain it can sever filaments and thus accelerate filament depolymerisation, a feature not seen in other formins (Chhabra and Higgs, 2006). Secondly this motif allows INF2 to bind to actin monomers that compete with DID, which results in activation of actin polymerisation, by preventing DID-DAD interactions (Ramabhadran *et al.*, 2013). The autoihibition via DID-DAD interactions inhibits the depolymerisation activity of INF2, but not the actin nucleation activity (Chhabra *et al.*, 2009).

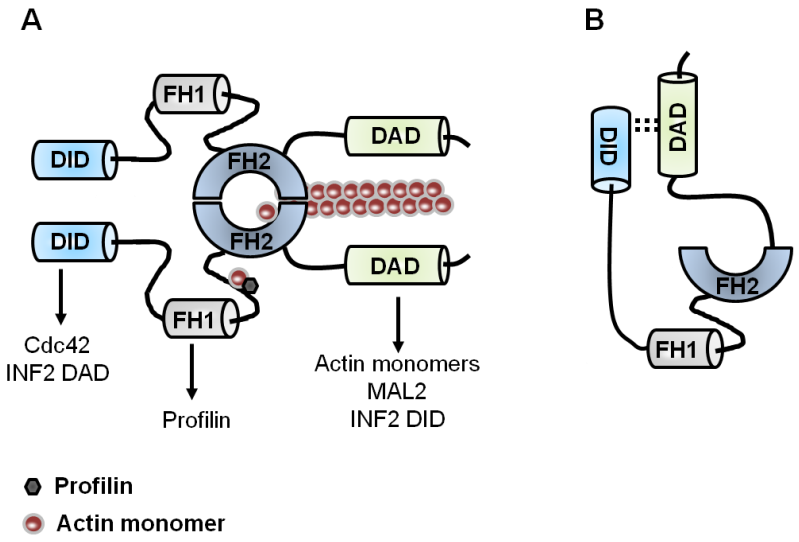

Figure 5-20 Functional INF2 dimer and autoinhibited structure.

A- Schematic diagram of an INF2 dimer interacting with an actin filament, where the FH2 domains form a ring-like structure around the actin filament (Gurel et al., 2014). Arrows indicate the interactions for the DID, FH1 and DAD domains. B- The autoinhibited INF2 conformation, due to DID/DAD interactions. Adapted from Campellone and Welch (2010).

There are two *INF2* C-terminal splice variants (CAAX and non-CAAX) that have different intracellular localisation, which suggests that each isoform might have a different function (Ramabhadran *et al.*, 2011). INF2 CAAX is farnesylated, enabling it to localise to the ER Golgi (Chhabra *et al.*, 2009). INF2 non-CAAX is found in the cytosol and thought to be involved in maintenance of the Golgi (Ramabhadran *et al.*, 2011). It has also been demonstrated that INF2 colocalised with proteins called myelin and lymphocyte protein (MAL), which are involved in apical transport of proteins (Puertollano and Alonso, 1999). This transcytosis transportation mechanism was shown to include the regulation of CD59 cell surface expression (Madrid *et al.*, 2010). This

was dependent on a Rho-GTPase (Cdc42) that bound to the N-terminus of INF2 and MAL2, which bound to the C-terminus of INF2 (Madrid *et al.*, 2010).

5.7.2. Disease association

INF2 has been shown to be expressed in Schwann cells and podocytes by Boyer *et al.* (2011b). This might explain why mutations in *INF2* have been previously associated with a glomerulopathy, Focal Segmental Glomerulosclerosis (FSGS) and a neuropathy, Charcot Marie Tooth (CMT) disease (Boyer *et al.*, 2011b). FSGS is characterised by proteinuria and podocytes foot process effacement (demonstrated by electron microscopy) (Bose and Cattran, 2014). Patients with CMT can experience muscle weakness and wasting, walking difficulties, a decrease in the number of myelinated fibres and reduced median-nerve velocities (Boyer *et al.*, 2011b).

Currently, published disease-causing mutations have occurred in patients who have either CMT and FSGS (Boyer *et al.*, 2011b, Mademan *et al.*, 2013, Rodriguez *et al.*, 2013, Toyota *et al.*, 2013, Park *et al.*, 2014) or FSGS alone (Barua *et al.*, 2013, Boyer *et al.*, 2011a, Brown *et al.*, 2010, Gbadegesin *et al.*, 2012, Laurin *et al.*, 2014, Lee *et al.*, 2011, Lipska *et al.*, 2013). Sequence variants found to date have been dominantly inherited and mainly occur in exons 2-4, encoding the DID domain, demonstrated in Figure 5-21. However there has been a report of a change that has occurred in exon 6 (Sanchez-Ares *et al.*, 2013), although this also codes for part of the DID domain. This highlighted the importance of the autoinhibitory function of DID domain.

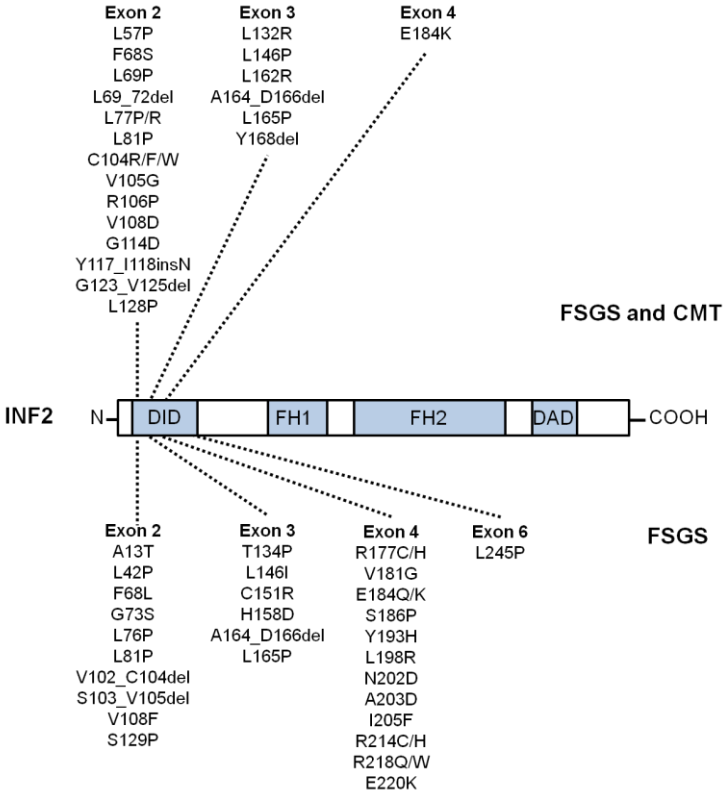

Figure 5-21 Reported mutations in INF2.

Top- mutations that were associated with CMT and FSGS. Bottom- mutations associated with FSGS only. Adapted from Mademan et al. (2013).

This section described the sequence variants found in *INF2* in this project.

5.7.3. Family 16

Family 16 had two members with aHUS and CMT disease. The proband presented with pex cavus and difficulty walking at age 7, leading to a diagnosis of CMT disease. She then presented with severe proteinuria at age 15, 5 days after a mild undiagnosed sore throat and after several months of malaise and fatigue. At presentation she was hypertensive (185/110mmHg), anaemic (Haemoglobin (Hb) 9.4g/dl), had a low platelet count (119×10^9/L) and had schistocytes present in a blood film. A daily regime of plasma exchange and haemodialysis was commenced. Treatment with eculizumab was started with continuation of haemodialysis. Platelet count increased temporarily (173×10^9/L), before falling to 100×10^9/L. Eculizumab concentration was sufficient and complement activity was completely suppressed (CH50). Three months later a kidney biopsy demonstrated TMA, with features of glomerulosclerosis, shown in Figure 5-22. After 9 months with no renal recovery eculizumab was withdrawn.

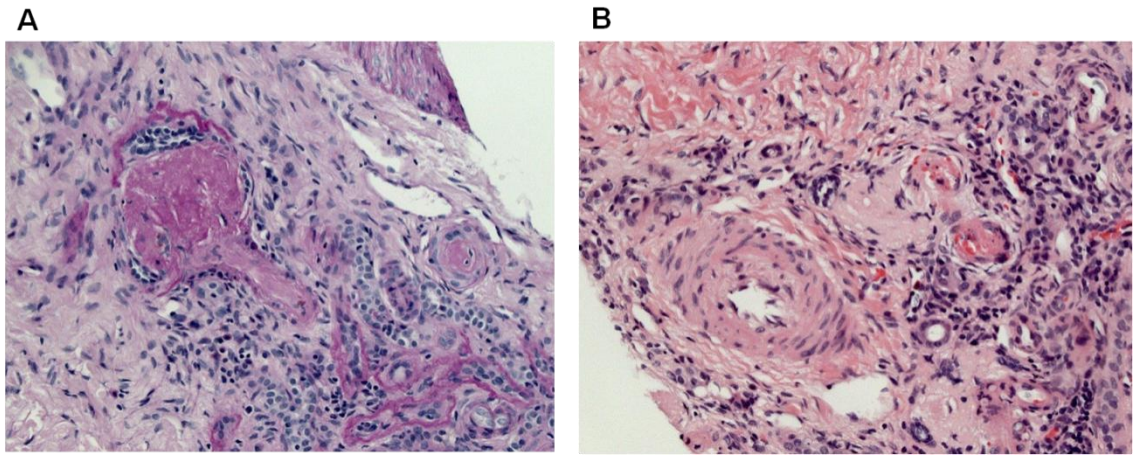

Figure 5-22 Renal biopsy of proband, family 16.

A- Glomerulus (left) showing global sclerosis and occluded arteriole (right), 20x periodic acid–Schiff (PAS) staining. B- two arterioles (right) showing features of active thrombosis and small artery (left) with some intimal oedema with fibrosis, 20x Haematoxylin and eosin (HE) staining.

The mother (I:2) presented aged 10yrs with walking difficulties and pex cavus. A motor axonal neuropathy was found with no sensory issues leading to a diagnosis of CMT. Age 17 she presented with proteinuria (12g/L) and an elevated serum creatinine. A renal biopsy demonstrated end-stage changes with diffuse global sclerosis. She had haemodialysis for 2 years when she had a renal transplant, which lasted for 6 years. After a further 5 years on dialysis she had another renal transplant. After 5 years she presented with worsening renal function, low platelets and low haptoglobins. A biopsy of the renal transplant demonstrated TMA, shown in Figure 5-23.

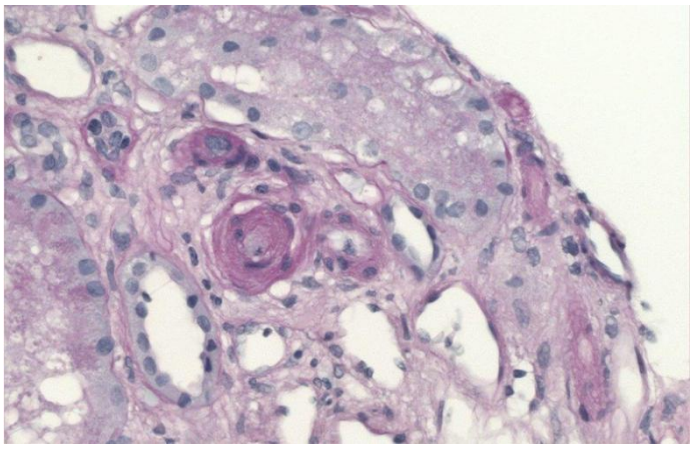

Figure 5-23 Biopsy of renal transplant from 1:2 of family 16.

Graft biopsy from the mother demonstrated an occluded arteriole (PAS staining).

The proband was found to have low C3 (0.56g/L, normal is 0.68-1.38g/L) and normal levels of C4 (0.19g/L), FH (0.39g/L) and FI (47mg/L). Testing for anti-FH

autoantibodies was negative. Routine MLPA screening found that they had 1 copy of *CFHR1* and two copies of *CFHR3*. They were screened for all known disease-associated genes (*CFH, CFI, CFB, MCP, C3, THBD* and *ADAMTS13*) and no pathogenic sequence variants were found. Genetic analysis did not detect the *C5* single nucleotide polymorphism (p.R885H) known to prevent eculizumab binding (Nishimura *et al.*, 2014). Testing for cobalamin C disorder was normal.

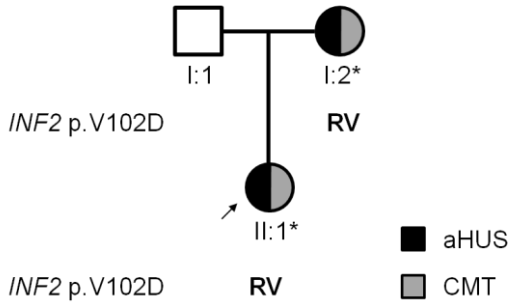

Figure 5-24 Pedigree of family 16.

Family members with CMT, are shown in grey and aHUS in black. Proband is indicated by the arrow and '' indicates all members tested by WES. Genotype was shown in bold.*

Due to the complex phenotype seen in this family, a gene list was created to see if there were sequence variants that may explain the observation of a renal and/or neurological phenotype. These gene lists can be seen in Table 45 and Table 46 of Appendix J. Screening of WES data for variants occurring in these genes, revealed a heterozygous change in exon 2 of *INF2*, c.305T>A, p.V102D. V102D was not reported in 1000g, ESP6500 or dbSNP. Autosomal dominant changes in *INF2* have been previously reported in association with both CMT and a renal disease, FSGS. Combined with the absence of this variant from control populations, *INF2* was considered a good candidate for disease. Sanger sequencing confirmed both affected individuals carried V102D in heterozygosity, the sequencing trace for the proband is shown in Figure 5-25.

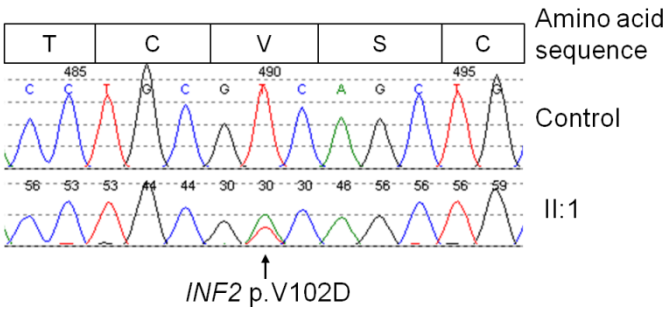

Figure 5-25 Sequencing chromatogram of V102D in family 16.

Sequencing chromatogram for control and proband (II:1), data for I:2 not shown. The amino acid sequence encoded by exon 2 is shown above.

In light of this finding, the candidate gene lists for the three inheritance patterns from section 5.5, were reviewed and revealed that in the dominant analysis, a second family had an *INF2* sequence variant, family 9.

5.7.4. Family 9

This family had two affected male offspring, both of whom presented with disease within the first year of life. The proband had recurrent disease, without developing ESRF. The proband (II:2) had low C4 (0.10g/L, normal range 0.18-0.6g/L) and normal C3 (1.18g/L), FH (0.72g/L) and FI (58mg/L) levels. MLPA indicated 0 copies of *CFHR1/3* and screening for anti-FH autoantibodies was negative. Panel screening of complement genes (*CFH*, *CFI* and *MCP*) revealed no sequence variants. The serum levels in the deceased brother (II:1) were normal for C3 (0.63g/L) and FI (53mg/L).

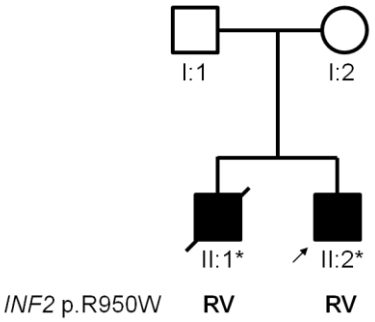

Figure 5-26 Pedigree for family 9.

Proband is indicated by the arrow and '' indicates all members tested by WES. Genotype was shown in bold.*

WES identified a sequence variant in exon 19 of *INF2*, c.2848C>T, p.R950W. It segregated in an autosomal dominant fashion and was reported in ESP6500 and 1000g

databases with MAFs of 0.04% and 0.05% respectively and was listed on dbSNP (rs199873407). Sanger sequencing confirmed that it segregated with disease, shown in Figure 5-27.

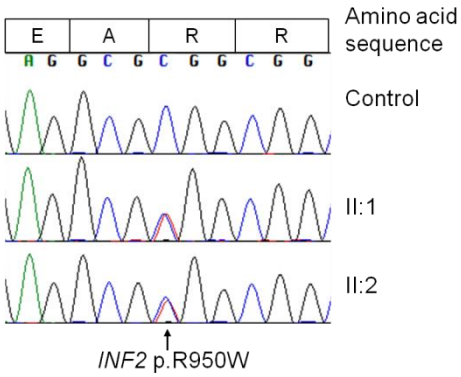

Figure 5-27 Sanger sequencing trace for family 9.

Sequencing chromatogram for control and affected patients (II:1 and II:2). The amino acid sequence encoded by exon 19 is shown above.

5.7.5. Sporadic patient screening

The finding of rare sequence variants in *INF2* in two familial cases, led to the hypothesis that there may also be *INF2* sequence variants within the sporadic aHUS population. In order to examine this, all coding exons of *INF2* were screened by Sanger sequencing in sporadic aHUS patients. Exons 2-4 contain the majority of disease-associated variants (Rodriguez *et al.*, 2013, Mademan *et al.*, 2013, Lee *et al.*, 2011, Brown *et al.*, 2010, Boyer *et al.*, 2011b, Boyer *et al.*, 2011a, Barua *et al.*, 2013) so were screened in 161 patients and the remaining exons 5- 22 were screened in 96 patients. However this did not yield any rare sequence variants. Subsequent analysis of exome data for 18 sporadic patients identified a rare heterozygous sequence variant in sporadic patient 4.

5.7.6. Sporadic patient 4

Sporadic patient 4, now referred to as Sp4, presented with aHUS at age 28 and developed ESRF. She had a living related transplant from her mother and there was delayed graft function due to intra- and post-operative hypotension. A biopsy at day 6 demonstrated features of acute tubular necrosis. By day 16 her creatinine had fallen to 220µmol/L. At day 23 she represented with hypertension, anaemia (Hb 6.0g/dl), thrombocytopenia (platelet count $96x10^9$/L) and a blood film demonstrating MAHA.

Her tacrolimus levels were always within the normal range. A renal transplant biopsy demonstrated TMA and plasma exchange was commenced. Over the next months she received daily plasma exchange with fresh frozen plasma and a further 2 transplant biopsies demonstrated on-going TMA. Her renal function continued to deteriorate and haemodialysis was initiated. Serum levels of complement proteins C3 (0.97g/L), C4 (0.26g/L), FH (0.50g/L) and FI (60mg/L) were normal. Anti-FH autoantibody screening and genetic screening (*CFH*, *CFI*, *MCP*, *C3*, *CFB* and *DGKE*) were negative.

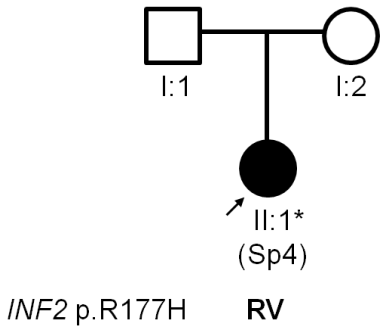

Figure 5-28 Pedigree for Sp4

Proband is indicated by an arrow and '' indicates all members tested by WES. Genotype was shown in bold.*

WES identified a heterozygous change in exon 4 of *INF2*, c.530G>A, p.R177H in Sp4, shown in Figure 5-29. This variant was not seen in 1000g, ESP6500 or dbSNP databases, however it was previously reported in patients with FSGS (Boyer *et al.*, 2011a, Gbadegesin *et al.*, 2012).

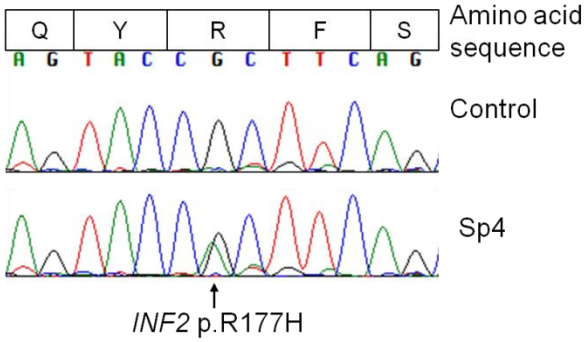

Figure 5-29 Sanger sequencing trace for Sp4.

Sequencing chromatogram for control and affected proband (Sp4). The amino acid sequence encoded by exon 4 is shown above.

Review of clinical data for this patient identified a family history of aHUS, however the proband will continue to be referred to as Sp4. The father (II:2), paternal uncle (II:3)

and cousin (III:2) all had a clinical diagnosis of aHUS. The father died at 25 years of age whilst on dialysis. III:2 presented age 36 with accelerated hypertension. A renal biopsy demonstrated very severe medial thickening and fibromuscular intimal proliferation with evidence of subacute arterial TMA. 8 glomeruli were sclerosed with a further two demonstrating segmental sclerosis. Serum levels of complement proteins C3, C4, FH and FI were described as normal. Anti-FH autoantibody screening and genetic screening (*CFH*, *CFI*, *MCP*, *C3*, *CFB* and *DGKE*) were negative. An updated family pedigree is shown in Figure 5-30.

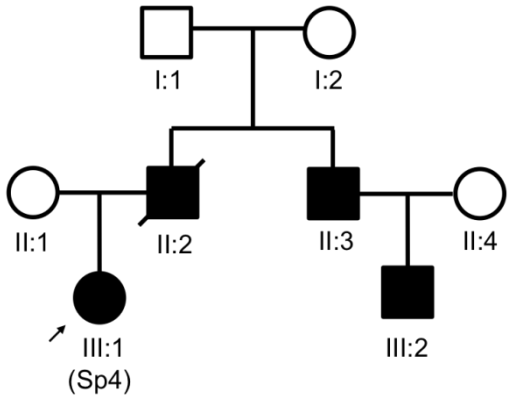

Figure 5-30 Updated pedigree for Sp4.

DNA was available for a cousin (III:2) and Sanger sequencing confirmed that he was also a carrier of *INF2,* c.530G>A, p.R177H.

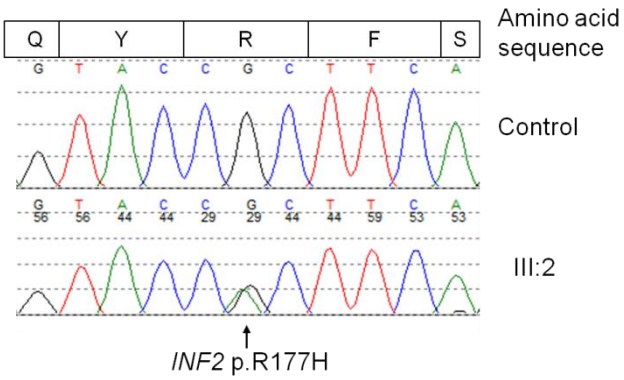

Figure 5-31 Sanger sequencing trace for the affected cousin (III:2) of Sp4.

Sequencing chromatogram for control and affected cousin (III:2) of Sp4. The amino acid sequence encoded by exon 4 is shown above.

The finding of 3 rare sequence variants, suggested that *INF2* was a good candidate for disease. They were found in two families and one sporadic case, which was later

demonstrated to be familial. Further analysis was performed to determine the predicted pathogenicity of the variants.

5.7.7. In silico analysis

To predict whether these three *INF2* variants would cause a detrimental effect to the protein, several *in silico* tools were used. All variants were described as deleterious by PolyPhen-2 (HDIV and HVAR) and Mutation Taster, shown in Table 27. V102D and R177H were also classed as deleterious by FATHM and Radial SVM, whereas R950W was predicted to be tolerated. Finally Mutation Assessor classed V102D and R177H as having a medium impact on protein structure and function, whilst R950W was described as neutral. These predictions suggest that V102D and R177H may perturb the protein to a greater degree than R950W.

Patient ID	Variant	*In silico* testing					
		PolyPhen-2		Mutation Taster	Mutation Assessor	FATHMM	Radial SVM
		HDIV	HVAR				
Family 16	V102D	D	D	D	M	D	D
Family 9	R950W	D	D	D	N	T	T
Sp4	R177H	D	D	D	M	D	D

Table 27 In silico predictions for the INF2 variants found.

D= deleterious, T= Tolerated, M= medium N= neutral.

The sequence conservation was then examined for each sequence variant by aligning the amino acid sequence for human, human with mutation and 10 other species. V102 was the least conserved, maintained only in chimp and orangutan. The presence of a valine in zebrafish was unlikely to be due to conservation, because surrounding amino acids are dissimilar to the human sequence. If the evolutionary tree is examined , the branch for the primates is after vertebrates such as Platypus and opossum, placental mammals such as mouse, rat, dolphin and dog. This suggested that the conversion to valine might have occurred within primates only. R177 is conserved across all available sequences and R950 is conserved in all available species apart from opossum and platypus.

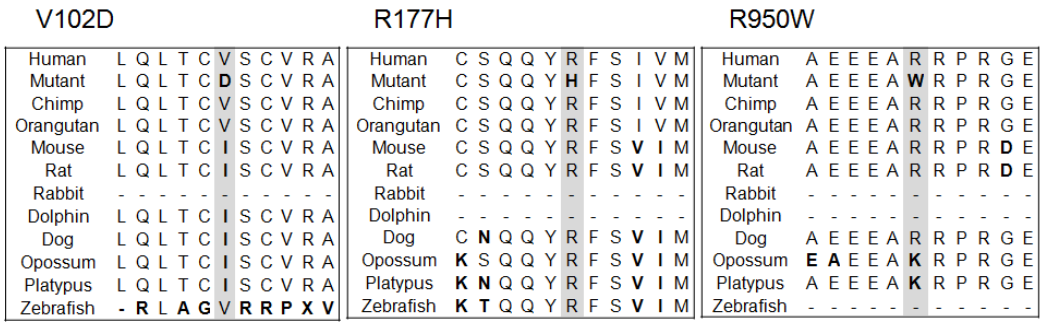

Figure 5-32 Protein sequence alignment of INF2.

Amino acid sequence alignment, showing the amino acid substitutions in INF2. Human with and without the variant are compared to 10 other species. All nucleotides in bold, represent differences to the human reference sequence.

The conservation scores calculated using GERP++ and PhyloP are shown below, in Table 28. Interestingly the GERP++ software described V102D as being most conserved, with a score of 4.76, compared to 4.48 and 2.53 for R177H and R950W respectively. This contrasts the results of the amino acid sequence comparison. Although it could be that V102 may be conserved across species that were not used in the alignment. R950W had the lowest conservation score using both programs, which suggested that it is in a region of INF2 that might be less critical to function. Indeed all mutations previously associated with CMT and FSGS are found in the DID domain, whilst little is known about the C-terminal portion of INF2.

Patient ID	Variant	GERP++	PhyloP
Family 16	V102D	4.76	1.777
Family 9	R950W	2.53	0.91
Sp4	R177H	4.48	2.027

Table 28 Conservation scores for INF2 variants.

5.7.8. Protein modelling

Two amino acid substitutions were found in the DID domain of INF2, an area with known functional importance. To examine the position of these variants within this domain, protein modelling was carried out. Phyre2 created a predicted protein structure that was manipulated in PyMOL. This would allow the visualisation of V102D and R177H in relation to the putatitive DAD binding site (amino acids R106, N110, A149, and I152). Figure 5-33 shows the predicted DID domain of INF2, annotated with the positions of V102D and R177H in relation to the amino acids involved in binding to the DAD. V102D was located close to the DAD binding site, particularly amino acid R106.

Although it was not present at the protein surface, it was predicted to disrupt the architecture of the 8th α-helix of the DID domain. R177H resided before the 13th α-helix of the DID domain and was surface exposed. However it was found on the opposite face of the DID domain, in relation to the DAD binding site.

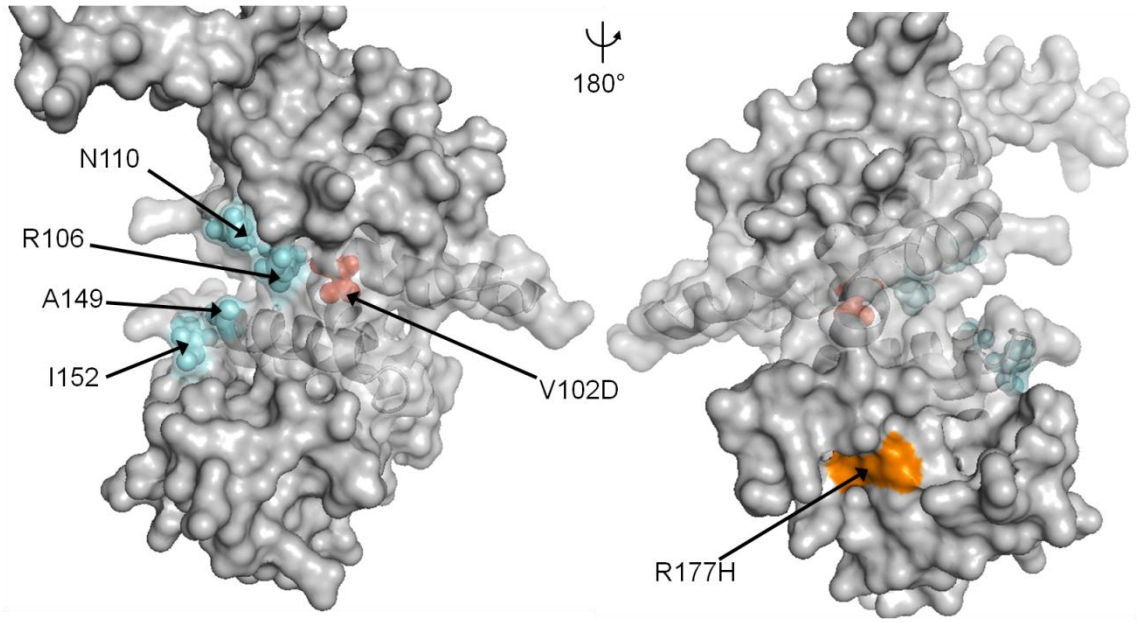

Figure 5-33 Protein model of INF2 DID domain.

Phyre2 was used to create a predicted protein model of INF2 DID domain, shown in grey. On the left is the DAD-binding site and on the right is the protein after a 180° rotation on y axis. Amino acids critical for DAD-binding (R106, N110, A149 and I152) are shown as blue spheres, V102D is shown as red spheres and R177H is shown in orange.

Figure 5-34A shows all reported mutations in the DID domain of INF2 associated with FSGS only (blue) or FSGS and CMT (yellow). This figure demonstrated that the mutations associated with FSGS and CMT or FSGS alone, cluster in distinct regions of the protein. Rose *et al.* (2005) had previously reported that mutations occurring in patients with FSGS and CMT, were found in the central portion of the domain, closer to the DAD binding site. Our results show similar results, with V102D and R177H segregating with the two respective phenotypes. Figure 5-34B shows the same information however showing the protein in surface format. This figure demonstrated that sequence variants associated with these two phenotypes, were found equally on the protein surface. It was observed that there may be more sequence variants, associated with a neurological phenotype, seen on the opposite face of the DID domain, in relation to the DAD- binding site. This might indicate a possible interaction site with a protein

found in the nervous system only, however there was no evidence in the literature to corroborate this observation.

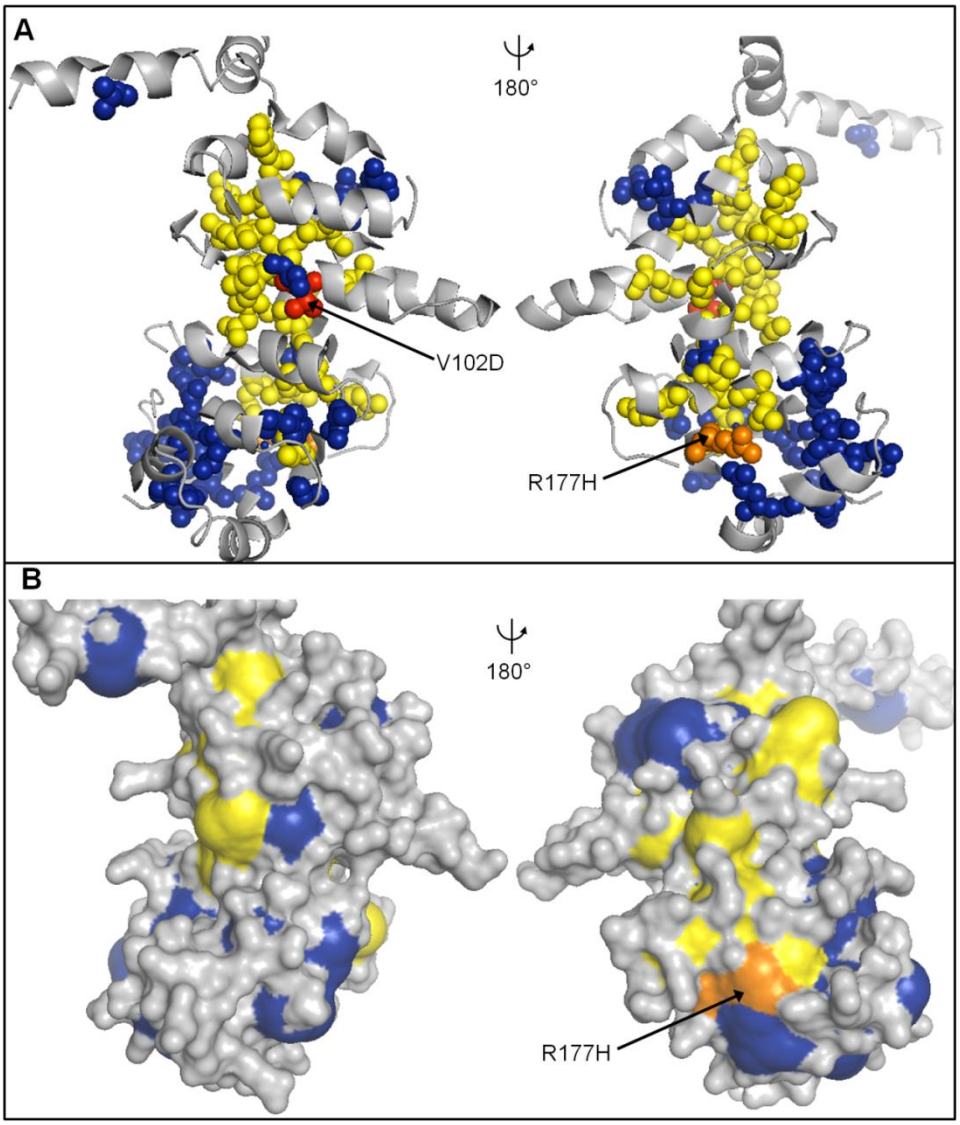

Figure 5-34 Reported mutations in INF2 DID domain.

Phyre2 model of INF2 DID domain showing mutations previously reported in the literature. A- Protein in a cartoon format. On the left is the DAD- binding surface and on the right is the protein rotated 180° on y axis. B- Protein in a cartoon format. On the left is the DAD- binding surface and on the right is the protein rotated 180° on y axis. Yellow- Mutations associated with CMT and FSGS. Blue- mutations associated with FSGS. Red- V102D. Orange- R177H.

These models suggested that V102D may have a greater effect on INF2 autoinhibition, as it was located in close proximity to the DAD binding site. Whereas R177H is located on the alternate face of the DID domain, clustering with other FSGS-associated sequence variants.

5.7.9. Pathophysiology

Podocytes are epithelial cells found in the glomerulus that are involved in formation of the GFB, along with the glomerular basement membrane (GBM) and fenestrated capillary endothelium (Swiatecka-Urban, 2013). Podocytes have complex structures that are dependent on the regulation of the cytoskeleton (Mathieson, 2012). Sequence variants in *ACTN4* (Kaplan *et al.*, 2000) *CD2AP* (Shih *et al.*, 1999) and *INF2* (Brown *et al.*, 2010) lead to alterations to the podocyte cytoskeleton, disrupting the GFB and have been found in association with nephrotic syndromes (Faul *et al.*, 2007).

The evidence demonstrated here indicated that *INF2* sequence variants V102D, R950W and R177H were rare and predicted to be pathogenic. V102D and R177H were found in the DID domain, an area containing sequence variants associated with CMT and FSGS or FSGS alone. R950W occurred closer to the C-terminus, between the FH2 and DAD domains, which had not undergone the same degree of functional analysis. This made it difficult to postulate how it is involved in disease pathogenesis, unless it prevented the protein from being secreted or disrupted the depolymerisation activity, known to be reliant on the C-terminal domains of INF2. In addition the clinical history of family 9 was very limited so it was unknown if there was a phenotype that would correspond to an *INF2* defect.

Functional analysis of R177H had been previously undertaken and was identified to disrupt INF2 cell localisation (Boyer *et al.*, 2011b). In control conditions INF2 had a perinuclear localisation, however when HeLa cells were transfected with an R177H mutant construct, INF2 was found in the cytoplasm (Boyer *et al.*, 2011b). These cells were found to have a disrupted actin cytoskeleton, which was thought to lead to disease formation by causing podocyte dysfunction (Boyer *et al.*, 2011a).

However *INF2* does not explain how these patients acquire aHUS. The occurrence of three rare *INF2* mutations, occurring in three separate aHUS cases, strongly suggested it was involved. INF2 disruption has been demonstrated to alter the intracellular transport of CD59 (Madrid *et al.*, 2010), therefore one possible mechanism for disease could be that these sequence variants altered the expression levels of the cell surface regulator CD59. Podocytes are important in maintaining the integrity of the GFB, in part due to the secretion of VEGF (Eremina *et al.*, 2008). It is well documented that VEGF inhibitors damage podocytes, predisposing patients to TMA (Eremina *et al.*, 2008).

Therefore the podocyte damage occurring as a result of FSGS could alter VEGF secretion and thus lead to a prothrombotic phenotype. As a result *INF2* may be a pleiotropic gene, causing both renal phenotypes (FSGS and aHUS).

Yet there have been no reports in the literature of patients with *INF2* mutations having TMAs. This included the 3 families with R177H mutations (Gbadegesin *et al.*, 2012, Boyer *et al.*, 2011a). There have been cases of patients with nephrotic disease who have later developed aHUS (Manenti *et al.*, 2013). This led to the hypothesis whether aHUS was actually a secondary phenomenon occurring after the FSGS-related damage, caused as a consequence of *INF2* sequence variants. Manenti *et al.* (2013) found that of 248 patients with the glomerulopathy, 6 developed aHUS. The underlying renal diseases were FSGS (1 case), MPGN (2 cases), C3 glomerulopathy (1 case) and small vessel vasculitis (2 cases). Then reviewing the literature they found a further 17 cases of nephrotic syndrome (FSGS- 8 cases, membranous glomerulonephritis (MGN)- 7 cases and minimal change disease- 2 cases), 17 cases of MPGN and 32 cases of vasculitis associated with aHUS.

There is a known risk haplotype in *CFH* and *MCP* that has been shown to increase an individual's risk of developing aHUS. The presence of additional risk SNPs may lead to increased complement activation and thus predisposing these patients to aHUS. Manenti *et al.* (2013) found that 4/6 patients with secondary TMA had *CFH*-H3 in homozygosity and 1/6 had MCP_{ggaac} in homozygosity with *CFH*-H3 in heterozygosity. It was hypothesised that the patients reported here might have these risk haplotypes. Therefore these patients were screened and the results are shown in Table 29 .

Risk SNPs	Genotype					
	Family 16		Family 9		Sp4 family	
	I:2	II:1	II:1	II:2	III:1 (Sp4)	III:2
CFH- H3	VV	VV	ND	RR	ND	ND
MCP_{GGAAC}	RV	VV	ND	RV	ND	ND

Table 29 Genotypes for Complement SNPs.

V- variant allele, R- reference allele and ND- not done.

The affected family members of family 16 were both homozygous for the *CFH*-H3 haplotype and the proband (II:1) was also homozygous for the *MCP* risk haplotype. Only one affected member of family 9 was tested (II:2) and they were found to be homozygous for the reference *CFH* haplotype and heterozygous for the *MCP* risk SNPs.

It would be interesting to see if either the second affected individual from family 9 (II:1) or Sp4 also carried these risk SNPs. If they do then it might suggest that additional genetic variants are necessary to cause aHUS in patients with *INF2* changes.

It is unusual that these patients also had disease recurrence in renal allografts as it suggests a circulating factor is responsible. Since INF2 is an intracellular molecule it is difficult to propose a mechanism in which it would lead to disease recurrence post transplantation. Sp4 had a living related transplant from her mother, however it was thought that the *INF2* variant was inherited from her father, who died at 25 years of age whilst on dialysis and who also had an affected brother and nephew. Therefore the possibility that the renal transplant also contained a genetic abnormality was low, although DNA testing of the mother was not carried out. Other studies have demonstrated FSGS recurrence post transplantation, including a patient with an *INF2* mutation (Gbadegesin *et al.*, 2012). However there are several other potential causes of TMA in a renal transplant including post operative infections (Murer *et al.*, 2000), anti-rejection drugs (Zarifian *et al.*, 1999, Ruggenenti, 2002) or acute rejection (Ponticelli and Banfi, 2006). Further work needs to be undertaken to determine the functional significance of these sequence variants and to try and elucidate a mechanism of disease in these individuals.

5.8. Complement component 9 (C9)

5.8.1. Gene

Complement component 9 (*C9*) is located on chromosome 5p13 (Abbott *et al.*, 1989) and encodes a 71kDa protein (Tegla *et al.*, 2011). C9 is amphipathic, meaning that it contains a hydrophobic C-terminus and a hydrophilic N-terminus that allows it to be positioned within lipid bilayers (DiScipio *et al.*, 1984). The complement system terminates in a common pathway that leads to the formation of MACs, of which C9 is a key component.

The first step in MAC formation requires the formation of C5b from the cleavage of C5 by a C5 convertase, produced from either the classical, lectin or alternative pathways (Walport, 2001a, Walport, 2001b). C5b can then bind to C6, followed by the addition of C7 and C8, shown in Figure 5-35A. At this point the complex is inserted into the cell membrane, due to the hydrophobic tail of C8 (Senior and Wallace, 2014). C9 then joins

the complex and self-polymerises to form a pore in the cell membrane, shown in Figure 5-35B. Approximately 6-16 C9 molecules have been shown to be involved in pore formation (Podack *et al.*, 1982). The formation of MAC is regulated by CD59, a membrane bound molecule that prevents C9 from interacting with the C5b-8 complex, or self-polymerising (Huang *et al.*, 2006).

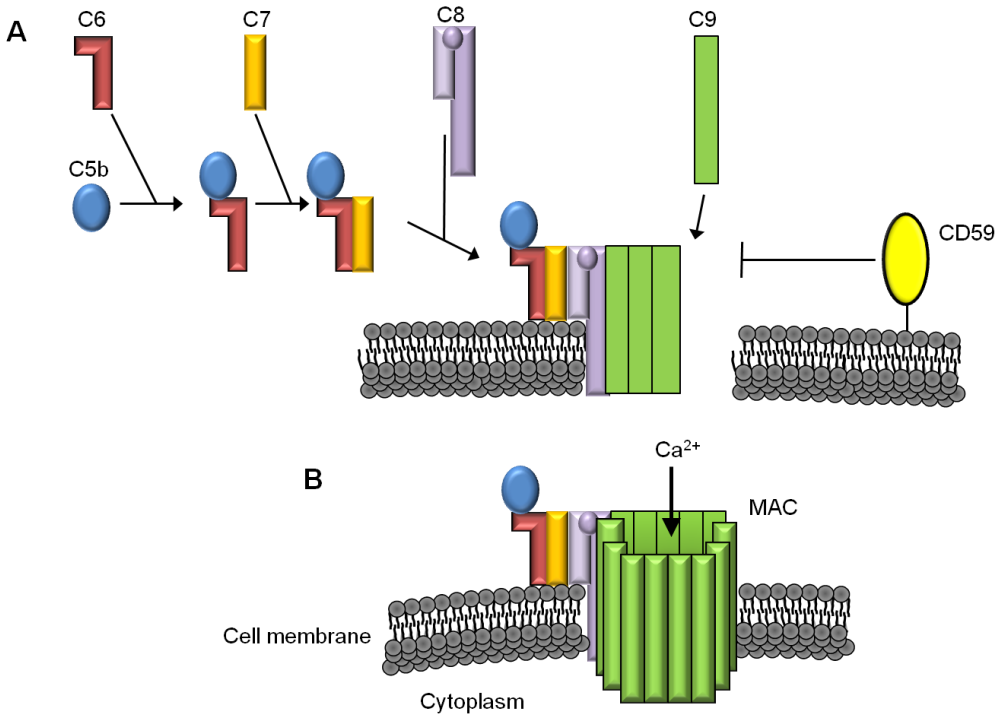

Figure 5-35 Schematic diagram showing MAC assembly.

A- Steps involved in MAC formation. C5b forms a complex with C6-8 and inserts into the cell membrane. C9 then binds and polymerises to form a pore structure.CD59, which inhibits MAC formation, is shown in yellow. B- The final structure of MAC. Figure is adapted from Bubeck (2014).

The use of the candidate variant pipeline to look for rare, high impact, conserved sequence variants that segregated with disease, identified a sequence variant in *C9* in one familial case. *C9* was on the candidate gene list due to its involvement in the complement system, which was known to be involved in the pathogenesis of aHUS. Therefore this sequence variant was selected for further analysis. This section described the sequence variants found in *C9* and how they were thought to cause disease.

5.8.2. Family 1

Family 1 has two affected siblings, both of whom have had disease recurrence after transplantation. The proband (II:1) has low C3 (0.63g/L) and normal C4 (0.29g/L), FI

(0.44g/L), and FI (68mg/L) levels. He had 2 copies of *CFHR1/3* and no anti-FH autoantibodies. Sanger sequencing of *C3*, *MCP*, *CFH*, *CFB*, *CFI* and *THBD* revealed no sequence variants.

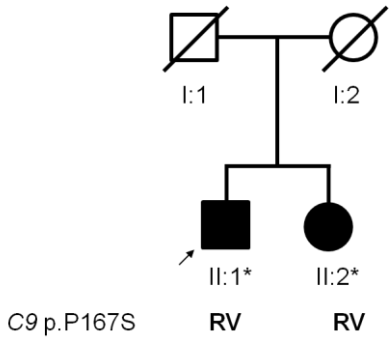

Figure 5-36 Pedigree of family 1.

Proband is indicated by the arrow and '' indicates all members tested by WES. Genotype was shown in bold.*

WES found a rare sequence variant in exon 5 of *C9*, c.499C>T, p.P167S. This variant had a MAF of 0.32% in 1000g, 0.6% in ESP6500 and was listed on dbSNP, rs34882957. Sanger sequencing confirmed that this variant segregated with both affected siblings, shown in Figure 5-37.

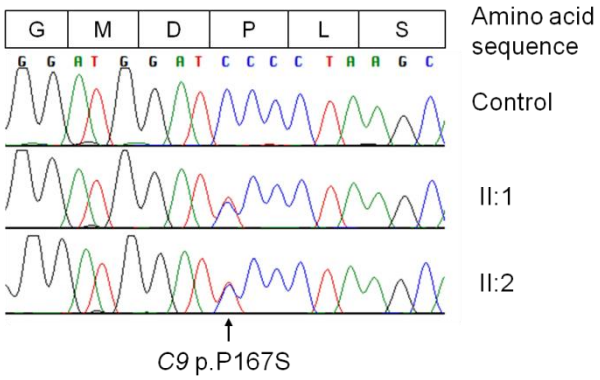

Figure 5-37 Sanger sequencing trace showing C9 P167S.

This variant was a good candidate as it was rare, segregated with disease and was found within the complement system, the major pathway involved in disease pathogenesis. It had also been previously associated with another complement-mediated disease, Age-related Macular Degeneration (AMD) (Seddon *et al.*, 2013). Finally it was found in a family with disease recurrence post transplantation, suggesting a serum factor, which is in keeping with a C9 abnormality. Therefore to investigate the possibility that there may

be additional patients with *C9* mutations, screening of sporadic aHUS cases was performed.

5.8.3. Sporadic patient screening

Firstly a panel of 96 sporadic aHUS patients were screened for exon 5 of *C9* by Sanger sequencing. This screening led to the discovery of another P167S sequence variant in a sporadic patient, referred to now as Sp5. Two other patients were found to have c.607A>G, p.I203V. This was listed on dbSNP (rs13361416) and had a MAF of 3% in 1000g and 2.7% in ESP6500. Due to the high MAF in both population databases, I203V was not further analysed. Secondly WES data for sporadic patients were reanalysed. This led to the discovery of *C9*, c.376G>A, p.G126R in a sporadic patient, referred to now as Sp6. This variant was listed on dbSNP (rs199939436) and reported in ESP6500 database with a MAF of 0.01%. Both variants were confirmed by Sanger sequencing, shown in Figure 5-38.

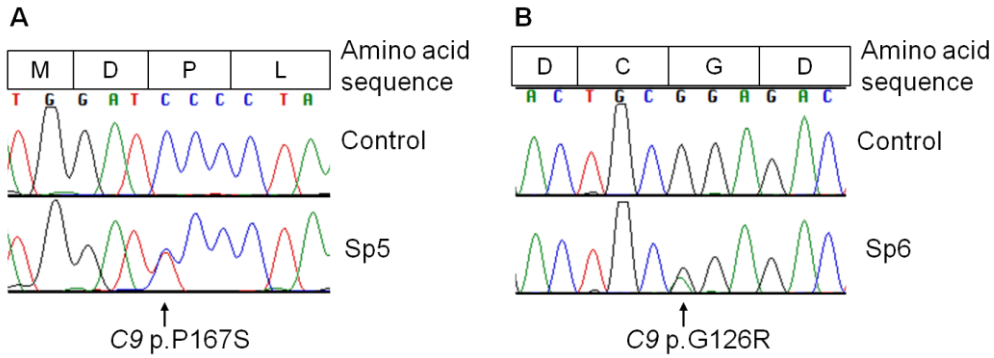

Figure 5-38 Sanger sequencing traces for C9 variants in sporadic cases.

A- Sequencing chromatogram for Sp5, confirming the P167S change. B- Sequencing chromatogram for Sp6, confirming the G126R change. The amino acid sequence encoded by the exon is shown above.

Sp5 presented at 15 years old with a rash and flu-like symptoms. She was positive for Measles virus and Parvovirus (serology was IgM positive). She had anaemia (Hb 6.1g/dL), low platelet count ($30x10^9$/L), MAHA (LDH 1489 IU/L) and no urine output. Complement serum levels were normal (C3 1.24g/L, C4 0.18g/L, FH 0.71g/L and FI 81mg/L). Disease rapidly resolved. Figure 5-39A shows the pedigree for Sp5.

Sp6 was a 38 year old female who presented with breathing difficulty from fluid overload, due to renal failure, requiring haemodialysis. She had no diarrhoea. Dialysis was discontinued after a year, due to some recovery of kidney function. 12 months later

a kidney biopsy showed chronic damage and her serum creatinine was elevated, which led to continuation of haemodialysis treatment. A year later she received a living related transplant from her unaffected brother. She had recurrent allograft HUS and despite plasma exchange, progressed to ESRF. It should be noted that the brother was not sequenced; therefore it was possible that he carried a sequence variant, which may be the cause of the disease recurrence. C3 and C4 levels were borderline low at 0.67g/L (normal 0.68-1.38 g/L) and 0.19g/L (normal 0.18-0.6g/L), respectively. FH and FI levels were normal at 0.46g/L and 62mg/L, respectively. MCP expression was normal and ADAMTS13 activity >10%. Routine genetic screening revealed that she had 2 copies *CFHR1/3* and no sequence variants in *CFH*, *CFI*, *MCP*, *CFB*, *C3* or *DGKE*. Figure 5-39B shows the pedigree for Sp6.

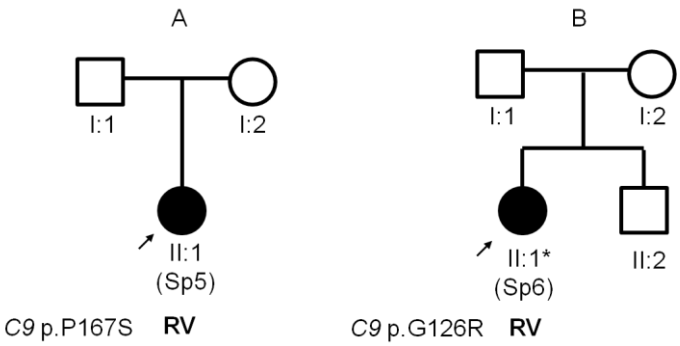

Figure 5-39 Pedigrees for Sp5 and Sp6.

A-Pedigree for Sp5 and B-Pedigree for Sp6. Proband was indicated by the arrow. '' indicated sample was sent for WES. Genotype was shown in bold.*

The variants were then analysed using *in silico* tools, to determine whether they may affect protein structure and function.

5.8.4. In silico analysis

The sequence variants found in *C9* were further analysed using *in silico* tools to predict whether or not a variant would be deleterious to the protein. The results are shown in Table 30. The two variants were predicted to be deleterious by PolyPhen-2 (HDIV and HVAR) and Mutation Taster. Neither variant was considered deleterious using Mutation Assessor, with a predicted impact of 'Medium'. FATHMM and RadialSVM classed P167S as tolerated and G126R as deleterious.

Patient ID	Variant	*In silico* testing					
		PolyPhen-2		Mutation Taster	Mutation Assessor	FATHMM	Radial SVM
		HDIV	HVAR				
Family 1, Sp5	P167S	D	D	D	M	T	T
Sp6	G126R	D	D	D	M	D	D

Table 30 In silico predictions for C9 variants.

D= deleterious, M= medium and T= tolerated.

The protein sequences were then compared to multiple species, in order to determine if they were evolutionary conserved areas. The results of the sequence alignment can be seen in Figure 5-40. The sequence for Zebrafish was not available for comparison. Both amino acid positions were well conserved, P167S was maintained through to the opossum and G126R to platypus.

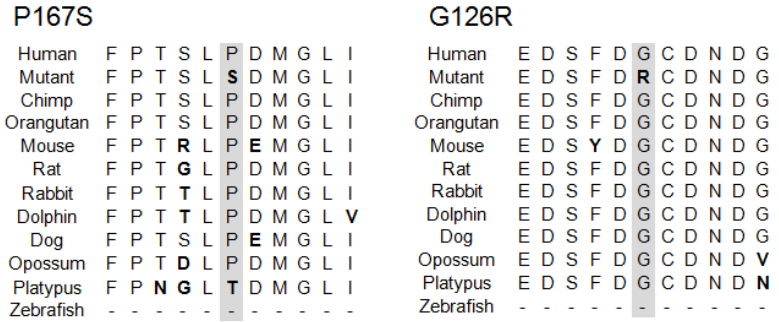

Figure 5-40 Protein sequence alignment of C9.

Amino acid sequence alignment showing the amino acid substitutions in C9. Human with and without the variant were compared to 10 other species. Grey indicates the position of either P167S or G126R. All nucleotides in bold, represent differences to the human reference sequence. Sequence not available in a species is represented with a '-'.

In silico predictions for sequence conservation were made using GERP++ and PhyloP. The scores obtained also demonstrated higher sequence conservation for G126R compared to P167S, however both were considered to be evolutionary conserved.

Patient ID	Variant	GERP++	PhyloP
Family 1, Sp5	P167S	3.81	1.401
Sp6	G126R	5.32	2.484

Table 31 Conservation scores for C9 variants.

To try and understand how these variants might affect C9 function, protein modelling was carried out.

5.8.5. Protein modelling

The predicted protein structure of C9 was calculated using Phyre2 and viewed using PyMOL software. Figure 5-41 shows the polymerisation surface of C9 either as a cartoon (A) or as a surface (B) model. On this protein model neither P167S nor G126R reside close to the CD59 binding site, amino acids 366-371 (Huang *et al.*, 2006). This suggested that these variants did not disrupt MAC regulation by CD59. One other hypothesis for their role in disease was that they improved the ability of C9 to polymerise, however the specific amino acids involved in polymerisation are not currently known. G126R was found on the protein surface so was thought to have a direct impact on C9 polymerisation, shown in Figure 5-41C. P167S was not present directly at the protein surface; however due to its close proximity, it was thought that it may also influence C9 polymerisation.

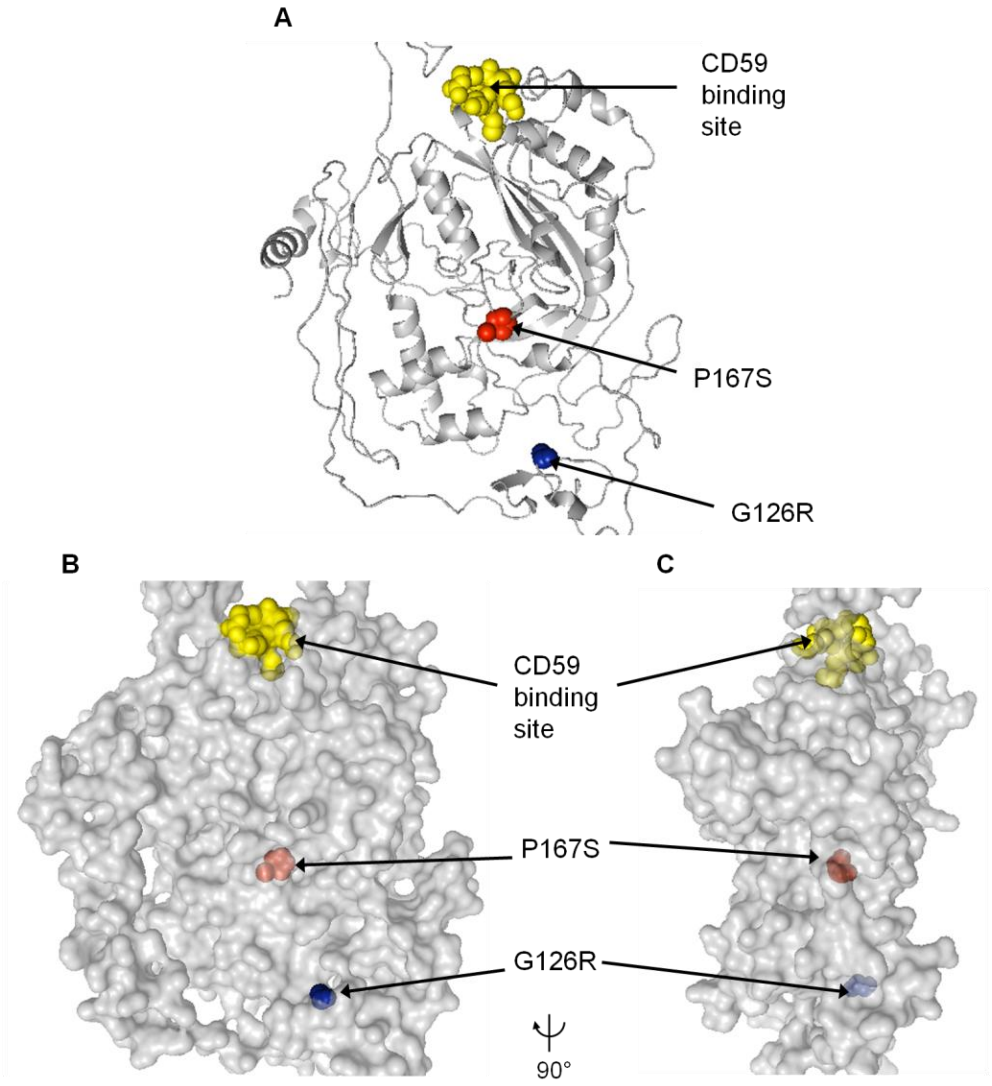

Figure 5-41 Protein model of C9.

Phyre2 predicted protein model of C9. A- The polymerisation surface protein in a cartoon format and B- polymerisation surface in a surface format, shown at transparency 0.4. C- Figure B was rotated at 90° on Y axis and shown at transparency 0.4. CD59 binding site was shown by yellow spheres, P167S is shown by red spheres and G126R is shown by blue spheres.

The evidence described here demonstrated that *C9* variants G126R and P167S were rare variants that were predicted to be deleterious. Modelling of mutations in the protein, suggested that they did not disrupt CD59 binding. Therefore an alternative hypothesis was that these variants enhanced MAC formation, by enhancing C9 polymerisation. The proposed mechanism through which these variants can cause aHUS, was described in the next section.

5.8.6. Mechanism of disease

There are several diseases that have involvement of the terminal complement pathway in disease pathogenesis. These include Paroxysmal Nocturnal Haemoglobinuria (PNH) (Takeda *et al.*, 1993) and AMD (Seddon *et al.*, 2013). Insertion of adequate quantities of MAC into a cell membrane, leads to cell lysis. When complement activation is reduced, sublytic amounts of MAC can initiate intracellular signalling mechanisms (Cole and Morgan, 2003, Bohana-Kashtan *et al.*, 2004). There has been a report of a patient with a *CFH* deficiency and low C9 levels (Falcão *et al.*, 2008). Low C9 levels may be the result of deregulated complement regulation and thus increased terminal complement activity, leading to enhanced consumption of C9. Currently it is unclear whether C5a, C5aR or MAC is predominantly important in aHUS. Identification of a functionally significant variant in the last component of the terminal pathway would support the importance of MAC formation in disease.

Witzel-Schlomp *et al.* (1997) first reported C9 deficiency as a result of compound heterozygous mutations in *C9*. These patients had recurrent bacterial meningitis (Zoppi *et al.*, 1990). It has since been shown that C9 deficiency is common in the Japanese population, with one homozygote occurring in every thousand individuals (Horiuchi *et al.*, 1998). There have been no reports of individuals developing TMA and it had been demonstrated that C9 deficiency conferred a reduced risk of developing AMD (Nishiguchi *et al.*, 2012). This is fitting as both diseases are thought to occur as a result of complement over activation. Therefore the mutations in *C9* reported here were predicted to be activating. This would result in a mutant MAC that had either increased formation or increased resistance to breakdown, to form what was termed here a ‘Super MAC’. This was predicted to have two possible effects; firstly increased cell lysis and secondly augmented sublytic signalling. The downstream effect of sublytic MAC signalling is shown in Figure 5-42 and reviewed below.

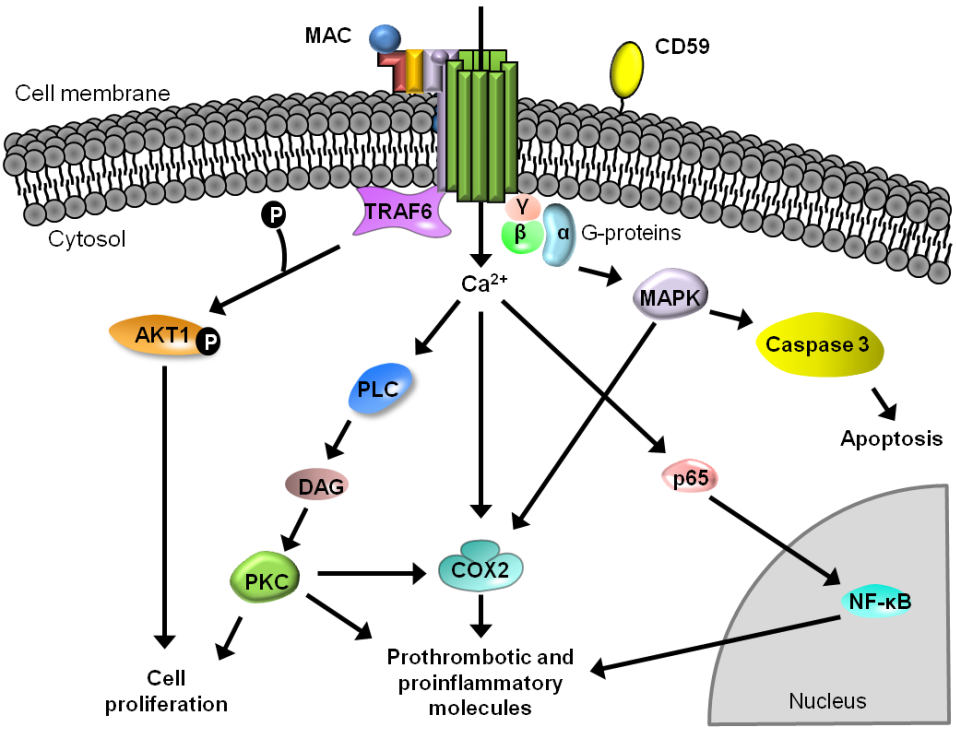

Figure 5-42 Schematic diagram of sublytic MAC signalling

Intracellular signalling pathways affected by MAC insertion into the cell membrane. Direct interaction of MAC with TRAF6 or G-proteins can lead to downstream pathways culminating in Cell proliferation or apoptosis. The influx of Ca^{2+} into the cytosol has several effects that can result in cell proliferation or the increased expression of prothrombotic or proinflammatory molecules. Adapted from Cole and Morgan (2003).

The PI signalling pathway has been demonstrated to be affected by MAC insertion. The loss of DGKε as previously described, led to increased PKC signalling, which was thought to mediate tissue injury (Lemaire *et al.*, 2013). Several studies have shown that the addition of MAC to rat glomerular epithelial cells also caused increased PKC signalling (Carney *et al.*, 1990, Cybulsky and Cyr, 1993). Therefore increased MAC sublytic signalling could cause tissue injury in a similar mechanism to DGKε functional deficiency. Some of these effects may be due to the rapid influx of Ca^{2+} into the cytosol (Carney *et al.*, 1990), causing increased PLC activation and thus more DAG formation (Cybulsky and Cyr, 1993). However some studies have shown that increased expression of DAG was also due to the direct effect of MAC, where the addition of EGTA (a chelating agent), still resulted in PLC activation and DAG formation (Cybulsky *et al.*, 1989). In addition the treatment of HUVECs, coated in MAC, with sphingosine (PKC inhibitor) was shown to inhibit the release of vWF multimers from α granules in platelets (Hattori *et al.*, 1989). This suggests that PI signalling might be up regulated in patients with *C9* mutations, similar to what is seen in DGKε deficient patients.

Qiu *et al.* (2012) showed that MAC induced mesangial expansion in rats with anti-Thy1 nephritis (a model for glomerulonephritis). The mechanism for this was based on MAC binding to TNF receptor-associated factor 6 (TRAF6) leading to the ubiquitination and phosphorylation of AKT1. This resulted in increased extracellular matrix (ECM) synthesis and proliferation of the mesangium. Additionally Halperin *et al.* (1993) demonstrated that MACs sensitize Swiss 3T3 cells to the mitogenic effects of platelet-derived growth factor (PDGF) and that MACs can also induce cell proliferation in the absence of exogenous growth factors. A study by Benzaquen *et al.* (1994) also showed an increase in expression of PDGF and basic fibroblast growth factor, potent mitogens, when rat mesangial cells were treated with MAC. Finally activation of PKC (Kraus and Fishelson, 2000) and phospholipase A_2 gamma ($PLA_2\gamma$) (Elimam *et al.*, 2013) by MAC has been shown to desensitize cells to complement-mediated lysis. As a result cells may continue to secrete prothrombotic or proinflammatory molecules. Increased cell proliferation may cause narrowing of the glomerular vasculature, potentially facilitating thrombus formation. However other studies have shown that MACs can lead to mesangial cell apoptosis (Nauta *et al.*, 2002). This was caused by MAPK signalling pathways, which up regulated caspase 3 expression (Nauta *et al.*, 2002, Cole and Morgan, 2003). It is possible MAC activates different intracellular signalling pathways, resulting in alternative downstream cellular responses.

Hansch *et al.* (1984) demonstrated that macrophages and monocytes treated with purified MAC components, led to the expression of proinflammatory molecules arachidonic acid, Prostaglandin E and thromboxane B2 and that these effects were independent of C3 and C5 cleavage products, C3a and C5a, which are known anaphylatoxins. They then replicated this study using platelets (Hansch *et al.*, 1985). This may be mediated by induction of COX2 in a MAPK, PKC or calcium ion-dependent mechanism (Takano *et al.*, 2001, Noris *et al.*, 2015). Other proinflammatory effects have been seen as a downstream effect of MAC formation. Lovett *et al.*(1987) found that treating mesangial cells with MAC increased interleukin (IL) 1β levels and Kilgore *et al.* (1997) reported the nuclear translocation of p65 in response to MAC, activating nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) signalling. Similarly the use of anti-CD59 antibodies on retinal pigmented epithelial cells saw the increase in secretion of IL- 6 and 8, monocyte chemoattractant protein-1, VEGF and matrix metalloproteinases (MMP) 2 and 9 (Lueck *et al.*, 2011). MMPs 2 and

9 may alter angiogenesis and ECM regulation (Stamenkovic, 2003, Bandyopadhyay and Rohrer, 2012), which could predispose the vasculature of the glomerulus to thromboses.

Prothrombotic effects of MAC include conversion of prothrombin into thrombin and increased platelet factor 4 release from platelets, which was not dependent on the lysis of cells (Wiedmer *et al.*, 1986). Hattori *et al.* (1989) incubated HUVECs with sublytic quantities of MAC components and found that it led to the release of platelet α granules. Importantly they found that the presence of C9 was critical for this to take place. The content of these granules include coagulation and growth factors, chemokines and adhesive proteins, which would promote thrombus formation (Nurden, 2011).

As stated previously there are two diseases associated with increased terminal complement activity, AMD and PNH. AMD is one of the most common causes of blindness in the elderly population, in the developed world (Hageman *et al.*, 2005). Risk of AMD has been associated with mutations in complement genes, including *CFH*, *CFI*, *C3* and *C9* (Seddon *et al.*, 2013, Hageman *et al.*, 2005). These mutations were hypothesised to lead to increased complement activity in the eye, leading to local inflammation (Klein *et al.*, 2005). Seddon *et al.* (2013) previously reported *C9* P167S in association with an increased risk to AMD, with an odds ratio of 2.2. Schramm *et al.* (2014) proposed that deficiency of C9 would be protective against AMD, whereas P167S would increase risk. This association was good evidence for the potential role of *C9* sequence variants in aHUS. Both aHUS and AMD are based upon complement deregulation, although the former is an acute disease and the latter is chronic. The tissues affected by both diseases, the glomerular basement membrane and Bruch's membrane, are similar in that they are exposed and fenestrated basement membranes, suggesting a common susceptibility to complement dysfunction (Kelly *et al.*, 2010).

PNH is a clonal disorder of erythrocytes that contain sequence variants in *PIGA*, which encodes the glycosylphosphatidylinositol (GPI) anchor, necessary for anchoring complement regulators to the cell membrane (Takeda *et al.*, 1993). It results in the loss of complement regulators CD59 and CD55 and leads to excess complement activation and thus augmented lysis of erythrocyte (Wong and Kavanagh, 2015). These patients have an increased susceptibility to thromboses. When CD59 is blocked in rats they develop glomerular thromboses, similar to PNH patients (Nangaku *et al.*, 1998). Eculizumab has been successfully used to treated patients with PNH, as it targets C5

preventing MAC formation and haemolysis (Hochsmann *et al.*, 2014, Hillmen *et al.*, 2006). There has been a case in the literature of a patient with PNH combined with C9 deficiency (Yonemura *et al.*, 1990). The patient had a milder phenotype, indicating the importance of C9 in haemolysis. It was proposed that the *C9* mutations reported here were not involved in CD59 binding; rather form a 'Super MAC' that caused increased cell lysis or increased sublytic signalling. It was also noted that there were no other rare variants, segregating with disease in genes encoding other components of MAC such as *C5*, *C6*, *C7* and *C8A/B/G*, with the hypothesis that they could also lead to enhanced MAC formation. Further analysis of these *C9* sequence variants is needed to determine whether or not these variants are functionally significant.

5.9. Other gene candidates in literature

There have been other gene candidates reported in the literature, to be associated with aHUS. This includes *CFB* (Goicoechea de Jorge *et al.*, 2007), *MMACHC* (Lerner-Ellis *et al.*, 2006) and *THBD* (Delvaeye *et al.*, 2009). Screening of familial cases for rare and segregating sequence variants in these genes was negative, as described in 4.7.

There have also been sequence variants were reported in Fukutin related protein (*FKRP)* in a Japanese familial case of aHUS (Watanabe *et al.*, 2014). These variants were inherited in an autosomal dominant pattern, with affected individuals having disease onset in infancy. Autosomal recessive *FKRP* mutations have previously been associated with congenital muscular dystrophy (Hewitt, 2009), where mutant FKRP was unable to process dystroglycan (Esapa *et al.*, 2002). The association of FKRP to a muscle phenotype suggested that the variants described by Watanabe *et al.* were less likely to be a cause of aHUS. Despite this, the families in this project were screened for sequence variants within *FKRP*. However this analysis yielded no rare candidate variants that segregated with disease.

Bu *et al.* (2014) reported several rare sequence variants in coagulation genes, within the American aHUS cohort. These genes included coagulation factor XII (*F12*), plasminogen (*PLG*) and *VWF*. However no rare sequence variants that segregated with disease were found in the families analysed in this project (shown in Supplementary data).

A homozygous family originating from North Africa was described to have aHUS, with onset in adolescence (Lemaire, 2015). WES analysis identified a homozygous change in all affected offspring in a gene *ST3GAL1*, p.G288D. ST3GAL1 is a sialyltransferase that regulates the expression of sialic acid-containing carbohydrates and is highly expressed in the kidney (Kitagawa and Paulson, 1994). The removal of sialic acid components has been demonstrated to lead to reduced FH binding (Ram *et al.*, 1998). The mechanism of disease in these patients was thought to be that loss of ST3GAL1's sialyltransferase activity would lead to reduced cell surface expression of sialic acid-containing carbohydrates. This would then reduce the amount of FH that can bind, decreasing complement regulation at the cell surface (Lemaire, 2015). Screening of exome data for all families in this study did not reveal any rare sequence variants in *ST3GAL1* that segregated with disease.

5.10. Families with no genetic candidates

There were 19 families that did not have a genetic cause for disease before analysis using the exome candidate pipeline. After the discovery of novel disease candidates *DGKE*, *INF2* and *C9*, the number of families with no known genetic cause of disease was reduced to 13. Therefore of the 28 families that were analysed in total, 54% had their genetic cause of disease identified. However this assumes that the novel genes are functionally demonstrated to be causative of disease, which has yet to be carried out.

There were several reasons why no genetic candidate could be found in all the families tested. In some cases there was a lack of clinical data, which meant that some family members may have been treated as affected when they were not, for example the father in family 7 did not have TTP. There is also the possibility that there was consanguinity that was not known or disclosed. To try and avoid this multiple analyses were carried out using each mode of inheritance (autosomal dominant, autosomal recessive and compound heterozygous). Detailed clinical data was also critical in helping to further reduce the number of candidate sequence variants. For example family 16 had a very detailed phenotype, consisting of a combined renal and neurological phenotype, which identified *INF2* as a good candidate for disease.

Many families within this project only had DNA available for testing in one individual. This makes the analysis more difficult, because the final number of variants after filtering is higher, shown in Table 20. In addition the majority of the samples sent for

WES were affected family members. In the future it might be beneficial to undertake WES in unaffected family members, to identify variants that are restricted to affected individuals.

For some samples the quality of the DNA was very poor, in some cases requiring WGA prior to WES. This has been demonstrated previously to lead to accurate sequencing (Hollegaard *et al.*, 2013, Hasmats *et al.*, 2014). However from the coverage data shown in Table 18, the percentage of the exome covered was much lower than in non-amplified samples. This may result in candidate sequence variants being missed.

There were also limitations to the technology used, for example WES only covers 1% of the genome; therefore disease-causing mutations outside of exome would not be detected. WES poorly sequences CG-rich or highly repetitive regions (Meynert *et al.*, 2014), which might result in loss of candidate sequence variants. In addition the capture step can introduce bias, preferentially capturing wild type alleles, which may lead to low coverage of candidate variants (Meynert *et al.*, 2014). Finally in this project INDELs were removed from the analysis, due to their low detection accuracy (Fang *et al.*, 2014). Therefore there is the possibility that this step has removed candidate variants, as was seen in family 13 for a sequence variant in *DGKE*.

5.11. Discussion

WES is a highly beneficial technology as it widens the search for genetic causes of disease. This is crucial as many patients do not present with 'classical' symptoms or in cases where diseases have overlapping phenotypes, patients might have been misdiagnosed. However it can generate large quantities of data that can make identification of disease candidates very difficult. In addition it only interrogates approximately 1% of genome; therefore it is possible that disease-causing variants have been missed.

This is a very rare disease, which meant that there were few large recessive pedigrees in the Newcastle cohort, available for testing. 1359 genes were identified with rare sequence variants that segregated with 1 family. This makes it very difficult to differentiate sequence variants that directly cause aHUS from those that do not. One method of overcoming this could be to combine the WES data from the Newcastle

cohort with the data of other international and European cohorts. A larger sample size may allow for the detection of novel gene candidates for aHUS.

The candidate pipeline used in this project was basic with scope for evolution, with the aim to discover the most obvious candidates for disease. In the future a more stringent approach to variant calling could be used. For example reducing the MAF threshold to 0.1% in 1000g and ESP6500 or keeping variants that are classed as deleterious by more than one prediction program. This would further reduce the final number of candidate variants in the analyses, aiding candidate gene discovery.

In silico tools were used to try and identify variants that may cause disease. They use various methods to assess the impact a variant may have on the protein structure and function; thus predicting whether or not it may be disease-causing. The detailed *in silico* results for known disease-causing variants, identified in this project, were shown in Appendix I and summarised below, in Table 32. In total, 6 variants occurring in the genes *MCP, CFI* and *ADAMTS13* had been previously reported to be pathogenic. 5 of these variants were missense and had predictions from 6 *in silico* tools, whilst one was splice site variant and had one prediction score. Overall, none of the variants were predicted to be pathogenic by all 6 programs. PolyPhen-2 HDIV predicted all 5 missense variants to be deleterious, Mutation Taster predicted 5/6 variants to be damaging and PolyPhen-2 predicted 4/5 variants to be detrimental to the protein. RadialSVM predicted 2 variants to be pathogenic, with Mutation Assessor and FATHMM only predicting one variant to be deleterious. The lack of agreement between the predictions of Mutation Assessor and FATHMM, with the *in vitro* data may suggest they were less accurate, although only a small number of variants were analysed. Overall these observations are similar to those found by Marinozzi *et al.* (2014), who previously used *in silico* tools to assess the pathogenicity of *CFB* sequence variants. They found that the predictions only corresponded with *in vitro* data for a third of variants analysed. However studies have shown that some *in silico* tools, such as PolyPhen-2, produce more accurate predictions for loss-of-function variants compared to gain-of-function (Leong *et al.,* 2015), which is the case of *CFB* variants. It still remains that to fully understand how these sequence variants may cause aHUS, further *in vitro* or *in vivo* analysis is required.

Gene	Sequence Variant	Reference	Deleterious *in silico* prediction scores
MCP	Y189D	Fremeaux-Bacchi *et al.* (2006)	4/6
	c.286+2T>G	Fremeaux-Bacchi *et al.* (2006)	1/1
CFI	I416L	Bienaime *et al.* (2009)	4/6
ADAMTS13	I673F	Matsumoto et al (2004)	3/6
	R1060W	Camilleri *et al.* (2008)	3/6
	R507Q	Veyradier *et al.* (2004)	3/6

Table 32 In silico predictions of known disease-causing variants.

This table shows sequence variants that have previously been shown to cause disease and the number of in silico tools that have predicted them to be deleterious. In silico tools assessed were PolyPhen-2 (HDIV and HVAR), Mutation Taster, Mutation Assessor, FATHMM and Radial SVM. An in silico prediction for a splice site change was only available using Mutation Taster.

In this project INDELs were actively selected against, due to detection inaccuracies. These can occur due to the PCR amplification of the library, particularly in regions of repetitive sequences, which can cause polymerase slippage (Krawitz *et al.*, 2010). They can also occurring when there is an error during read alignment to the reference genome (Li, 2014). In the future it would be important to reanalyse data, specifically looking at INDELs. This is particularly true as technology improves and bioinformatic tools become more sensitive, rerunning samples may lead to more accurate variant calls, aiding discovery of pathogenic mutations.

The occurrence of CNVs was investigated in the RCA cluster, however there is the possibility that they have occurred elsewhere in the genome. It is possible to utilise WES data to look for CNVs, by comparing the read depth between a sample and an unrelated control (Plagnol *et al.*, 2012). This was not investigated here, but would be important to consider in the remaining families with no genetic cause of disease. Alternatively array comparative genomic hybridisation (CGH) could be carried out.

Overall WES identified 3 genes that are predicted to be novel genetic causes for disease, based on *in silico* analysis. Firstly sequence variants were identified in *DGKE*. This is involved in the regulation of the PI signalling pathway. Sequence variants were identified in homozygosity or compound heterozygosity, suggesting complete absence of functional DGKε was necessary to lead to disease onset. Patients were noted to develop disease in infancy and had disease recurrence whilst on eculizumab treatment. Other studies have identified similar findings, supporting the hypothesis that this was a

disease-associated gene. Loss of DGKε was hypothesised to cause increased PKC signalling, leading to the formation of a prothrombotic environment in the glomerulus. However the exact mechanism of disease has yet to be elucidated and functional validation of the variants identified is required.

The second novel genetic cause of disease was *INF2*. Several patients were identified to have sequence variants in this gene, which has been associated with renal and neurological pathologies. INF2 is involved in actin cytoskeletal regulation and genetic defects have been demonstrated to disrupt this, leading to podocyte dysfunction. R177H had previously been demonstrated to be functionally significant and V102D was positioned close to the DAD binding site so was hypothesised to be pathogenic. This was supported by the observation of a neurological phenotype in the affected family members. The combination of limited clinical information and lack of functional data meant that the pathogenicity variant R950W was unknown. One hypothesis is that *INF2* may be a pleiotropic gene, causing both renal phenotypes (FSGS and aHUS). Although cases of FSGS have been seen in association with aHUS, no mutations in *INF2* have been reported in patients with the concomitant features. Admittedly the relatively recent discovery of *INF2* mutations in FSGS may account for this. An alternative hypothesis is that aHUS may be occurring as a secondary phenomenon. The kidney damage caused by the *INF2* defect in addition to complement risk SNPs was proposed to predispose these patients to a TMA. However there are other post-surgical complications that can lead to TMA post transplantation, which have to be considered. Functional work is required to determine the role of these sequence variants in disease pathogenesis.

Finally rare sequence variants were identified in *C9*. These were hypothesised to lead to the formation of a 'super MAC' as a result of either enhanced C9 polymerisation or increased resistance to breakdown. The exact mechanism has yet to be elucidated in the context of aHUS; however the combination of these effects could create a prothrombotic environment in the glomerulus. Due to the association of the variant P167S with AMD, also mediated by complement over activation, it was hypothesised that it could play a role in the pathogenesis of aHUS. Whether P167S has a direct effect on disease onset or whether the presence of additional genetic factors are responsible for disease, has yet to be elucidated. Further interrogation of the WES data is needed to look for other sequence variants that may influence disease pathogenesis. The aHUS cohort could be screened for the remaining *C9* exons, which may identify additional

candidate variants. The *C9* sequence variants identified here require functional analysis to establish whether or not they are causative of aHUS.

Overall there were 13 families remaining with no genetic candidate for disease. This may be due to limited clinical information, poor quality DNA, filtering out INDELs, the occurrence of a pathogenic sequence outside of the exome or problems with the technology. There had been other novel gene candidates reported in the literature as being associated with aHUS. This included *FKRP*, *ST3GAL1* and coagulation genes *F12, PLG,* and *VWF*. However in this project no rare variants that segregated with disease were found in these patients.

Chapter 6: Discussion and Future work

6.1. Summary

The aim of this project was to elucidate the genetic cause of disease in aHUS patients, in whom no genetic cause had previously been identified.

6.2. Hybrid *CFH/CFHR3* gene causes aHUS

A patient was found to have a deletion of the C-terminal exons of *CFH*, which was hypothesised to have occurred as a result of MMEJ. This led to aberrant splicing of *CFH*, where exons of *CFHR3* were incorporated. This was confirmed in patient gDNA and evident in cDNA. Analysis of the serum identified a hybrid protein and a hybrid degradation product. Mass spectrometry confirmed the hybrid protein contained peptides that were identical to the FH/FHR3 reference sequence. Functional analysis of the patient purified FH mixture produced results that were consistent with a cell surface-binding defect. This was predicted to be the result of the loss of the C-terminal domains of FH, which prevented FH from binding to GAGs on the cell surface. However it may also be in part due to the propensity for the protein to break down, as demonstrated by western blotting.

6.3. High prevalence of genetic abnormalities in *CFH* and *CFHRs*

Studies have identified that *CFH* is the most frequently observed gene defect, occurring in approximately 30% of aHUS cases (Kavanagh *et al.*, 2013, Maga *et al.*, 2010, Fremeaux-Bacchi *et al.*, 2013, Noris *et al.*, 2010). Review of all patients in the Newcastle aHUS cohort with a rare *CFH* defect (n=95), identified 36.8% had a genomic rearrangement or conversion between *CFH* and a *CFHR*. This most commonly occurred between *CFH* and *CFHR1*, which was thought to be due to the high sequence homology shared between the two genes. Gene rearrangements were the result of NAHR, with the exception of two hybrid genes occurring between *CFH* and *CFHR3*, which were the result of MMEJ. The high prevalence of gene conversions and rearrangements in this gene cluster, emphasised the importance of undertaking CNV analysis in aHUS patients, as gene rearrangements were not detected by Sanger sequencing.

Approximately 41% of point mutations and INDELs were located in CCPs19-20 of *CFH*, an area critical for cell surface binding (Ferreira *et al.*, 2009). This demonstrated

the importance of complement regulation at the cell surface in preventing onset of aHUS. Sequence variants occurring in CCPs 1-18 were high impact mutations, with 57.7% comprising of nonsense, INDELs or affecting splice sites. Finally the majority of the sequence variants observed were heterozygous (96.8%), supporting previous evidence of *CFH* haploinsufficiency (Kavanagh *et al.*, 2013, Maga *et al.*, 2010, Fremeaux-Bacchi *et al.*, 2013, Noris *et al.*, 2010).

6.4. Known genetic causes of TMA were identified

28 familial cases with no known genetic cause of disease were analysed using a combination of Sanger sequencing and WES. In total 7 families had sequence variants in genes known to be associated with aHUS, which were *CFH*, *MCP* and *CFI*. All variants were observed in heterozygosity except for one family with a homozygous *MCP* sequence variant. All variants were rare and predicted to be deleterious by *in silico* tools.

Two families were demonstrated to have compound heterozygous mutations in *ADAMTS13* that were known to be functionally significant. Further detailed discussion with clinicians looking after these families, revealed a clinical phenotype in keeping TTP. The screening of familial aHUS cases yielded no rare sequence variants that segregated with disease in *CFB*, although rare, functionally significant sequence variants have been identified in the sporadic cohort. Screening of *THBD* in familial and sporadic cases did not identify rare variants that segregated with disease. Therefore no evidence was found in this cohort that supported *THBD* as being associated with disease. Finally no sequence variants were identified in *MMACHC* in either the familial or sporadic cohort, which might be due to the distinctive phenotype, which means that clinicians are not referring these patients to the national aHUS service.

6.5. Novel gene candidates identified as disease-causing

WES was used to look for novel genetic causes of disease in the remaining 19 families, with an unknown genetic aetiology. A filtering pipeline was used to reduce the number of candidate variants, in order to identify the most obvious candidates for disease. However there were caveats that may have led to the removal of candidate sequence variants.

DGKE was the first novel candidate gene proposed in this project. It was observed in multiple pedigrees, inherited in an autosomal recessive and compound heterozygous mode of inheritance. All variants were very rare and predicted to be deleterious by *in silico* tools. Loss of DGKε functional activity was proposed to lead to increased PKC signalling which would lead to formation of a prothrombotic environment and thus onset of disease. This gene has since been identified as a candidate gene for disease in other cohorts, supporting the findings reported here. Further functional work is required to demonstrate whether the variants identified in this study contribute to disease.

INF2 has previously been demonstrated to cause FSGS with CMT or FSGS alone. Rare variants in *INF2*, which were predicted to be deleterious, were found to segregate in 3 families with aHUS. Two variants were located in a mutation hot spot, one of which has been demonstrated to be functionally significant. The third variant arose in another region of the protein, which has not been functionally described and we cannot be definitive about its functional significance. This could be an example of genetic pleiotropy, where mutations in a single gene may lead to different phenotypes. However there is no evidence in the literature of patients with *INF2* changes having a TMA phenotype. It was then hypothesised that the TMA was a secondary event occurring after the FSGS-related kidney damage, potentially due to concomitant presence of *CFH* and *MCP* risk haplotypes. However it is possible that aHUS was the result of post-transplant milieu. Further analysis of the variants identified is required to establish whether they are causative of aHUS.

Sequence variants were identified in *C9* in 1 familial and 2 sporadic cases. All variants were rare and predicted to be deleterious by *in silico* tools. These variants were hypothesised to cause enhanced MAC formation, leading to increased cell lysis or increased sublytic signalling. One variant was previously associated as a risk variant in AMD, another disease caused by complement over activation. No other rare variants were seen in other components of MAC, such as *C5*, *C6*, *C7* and *C8A/B/G*, with the hypothesis that they could also lead to enhanced MAC formation. However it is yet to be established as to whether these variants are involved in disease pathogenesis.

6.6. Remaining families with an unknown genetic aetiology

The genetic analysis carried out in this project did not identify candidate sequence variants in all families tested. There may several reasons why a candidate gene was not

identified, firstly a causal intronic variant would not be detected by WES. Secondly families are referred to the National aHUS service as having a clinical diagnosis of aHUS; however the lack of clinical data available may lead to the possibility that there are additional phenotypes that may indicate an alternative diagnosis. There are also technical issues with the technology used that may lead to loss of candidate variants, for example low sequencing coverage of regions that are GC-rich or contain repetitive sequences. Finally, screening of novel genes reported in the literature as being associated with aHUS, such as coagulation genes (*F12*, *PLG* and *VWF*) (Bu *et al.*, 2014), *FKRP* (Watanabe *et al.*, 2014) and *ST3GAL1* (Lemaire, 2015), revealed no rare sequence variants that segregated with disease.

6.7. Final remarks

In this project 28 families with an unknown genetic aetiology were analysed and in total, 54% had their genetic cause of disease identified. However this assumes that all families in whom a novel candidate gene was identified, are functionally significant. The overall percentage of familial aHUS cases with no genetic cause of disease was reduced from 55% in 2010, before the project commenced, to 27%, after completion of the project, shown in Figure 6-1. The discovery of novel disease candidates *DGKE*, *INF2* and *C9*, comprised of 11% of the cohort, although this is subject to functional validation. Known TMA-associated gene *ADAMTS13* was observed in 4% of the cohort.

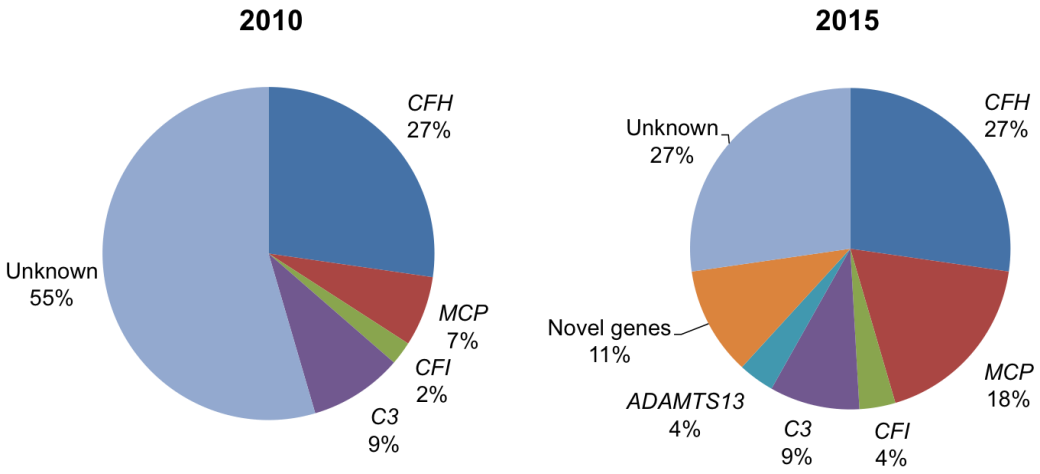

Figure 6-1 Genetic cause of disease in 2010 and 2015.

Pie charts showing the genetic cause of disease in familial aHUS cohort. The results are shown for the cohort in 2010 (before) and 2015 (after) completion of this project. This assumes that all sequence variants identified in novel genes are confirmed to be functionally significant.

6.8. Future work

In Chapter 3 functional analysis of the hybrid FH/FHR3 protein was performed using a haemolytic assay. This was carried out using OX24-purified FH species from control and patient. If more patient serum could be obtained, it would be interesting to perform the haemolytic assay using pure FH/FHR3 protein, rather than using a mixture of wild type and hybrid protein. The FH/FHR3 protein could be isolated using immunoaffinity chromatography, using the antibody specific to the *CFH* 402Y SNP, which was shown to be specific to the FH/FHR3 protein.

Patients with sequence variants in known disease-causing genes, described in chapter 4, should have functional analysis performed to determine their effect on protein. This includes testing for changes to protein expression in patient samples. The function of mutations could also be tested *in vitro*, for example using techniques such as surface plasmon resonance to detect alterations in binding affinity and haemolytic assays to detect changes to complement regulation at the cell surface.

Patients with *DGKE* sequence variants could be tested for absence of DGKε expression. For example Lemaire *et al.* (2013) used an immunohistochemistry method to stain kidney tissue with antibodies specific to DGKε. If kidney tissue was available from affected patients in this project, it would be interesting to examine any changes in DGKε staining in endothelial cells and podocytes. Further work should also be

undertaken to try and elucidate the mechanism in which DGKε deficiency causes aHUS and the functional significance of the sequence variants identified here.

Three *INF2* sequence variants observed in this project, one of which has been functionally analysed (R177H) (Boyer *et al.*, 2011b). It would be important to perform functional analysis on V102D and R950W to test their ability to regulate the actin cytoskeleton and to see whether they have a comparable defect to R177H. This is particularly true for R950W, which occurs in a different region of the protein. In addition it is important to try and determine the mechanism that causes TMA and disease recurrence post transplantation in these patients.

Sequence variants were observed in *C9* that were predicted to lead to either enhanced MAC formation or increased sublytic signalling. An aHUS panel was screened for exon 5 of *C9*; this should be extended to all other exons, as it may lead to the identification of further candidate sequence variants. It would be interesting to see if these patients have altered terminal complement activity. This could be tested for using patient sera, isolating patient C9 and testing it in a haemolytic assay. This would confirm if the mutant MAC alters the amount of cell lysis. Recombinantly expressed mutant C9 could be tested *in vitro* on a glomerular endothelial cell line to see if mutant MAC alters sublytic signalling, which may modify expression of prothrombotic and proinflammatory molecules or cell survival. Finally, it would be interesting to examine the remaining aHUS cohort for rare variants occurring in other components of MAC, such as C5, C6, C7 and C8, with the hypothesis that they also lead to enhance MAC formation.

The remaining families with no genetic candidate for disease could be tested for array comparative genomic hybridization (CGH) or WGS, to look for potential intronic causal variants. However initially, further interrogation of the WES data is necessary to look for other sequence variants that may mediate disease pathogenesis. This includes analysis of INDELs and CNVs, which was not carried out in this project. The analysis of these families might become easier as more families with aHUS are recruited, increasing the sample size.

Chapter 7: Appendices

7.1. Appendix A

The following figures were reprinted from Challis *et al.* (2015), with permission from the publisher. The manuscript can be seen in Appendix J.

Figure 3-2 Histological sections of patient renal biopsy.

Figure 3-3 MLPA screening results

Figure 3-4 Sequencing chromatogram of break point.

Figure 3-5 Breakpoint sequence homology.

Figure 3-8 Confirmation of CFH/CFHR3 cDNA product.

Figure 3-10 Binding sites for anti-FH antibodies.

Figure 3-11 Western blotting results of sera from proband and control.

Figure 3-12 Mass spectrometry results.

7.2. Appendix B

7.2.1. In house Sanger sequencing primer sequences

Gene	Exon	Forward Sequence (5' → 3')	Reverse Sequence (5' → 3')	Annealing temperature (°C)	Product size (bp)
ADAMTS13	1	CCCTGAACTGCAACCATCTT	CAGCTAGCTCAGAATA	55	316
	6	CTGTTTTCTCTCACCGAGG	TCAGCTCCTTGCCAGATC	55	210
	12	TACTGTAGGGTGCCATGTA	CCAGAGCCTGAACCACTTTG	55	336
	13	CCCAGATGCAAAGGATGAAG	ACTGCAGCTCCATCA	55	341
	17	TGGCCATGACAGTGACCCT	GTGCTGTGGATGAGCCTTC	55	228
	21	CACACACGCCACTTCCTG	CCACGTGTTCCCATATAGTCTG	55	360
	24	GGCTCAGTGGCTGCACTTTCC	TCCAGCGTCCCCAAACCTAAG	55	576
	29	AAGGCTTCCGTGAGTGCTA	AAGCCAGGTCTGGACAGCTTA	55	538
C9	4	GATACCTCACCTCCAGGGTTA	CACCTATGTCCCTCGCACAAA	60	352
	5	GTGTTGTGGTATTTCCACTTCTG	TCCAAACTACATCGCCTCTTC	60	330
MCP	5	TTGTGTACACCACCTCCA	TTATACAGGAGGAGGAAGCA	55	306
	2	GAAGCTATGGAGCTCATTGG	ATGCATTGTTATCTTGAACCTC	55	289
CFH	12	CTCACTTTATTGTGGCATATGT	CATGTATTGAGTACTGAATACCTGAG	55	417
DGKE	c.888+40A>G	CAGATGAAGGAAAATGTTGGA	GGATGGCTTGAACCTGAGAA	60	398
INF2	2	TGGTGGCCAGGAGGACA	CTGAAGCCTCATACCAGGTC	55	524
	3/4	TGCACAGGCATGGGAAG	ATGCACTGGCCAAGAGGC	55	466
	5.1	CTCTGGGTGGCAGAAGTGA	CCTTGTTCTAGTGGGCCTG	63	570
	5.2	GGACGCAAGATGGAGACA	ACCTGAATGAACTGCATG	55	550
	6	TGGTACACTGGCGCTGAC	CCCAGCCTCTGGACCTCA	60	300
	7	CACAGTCCTTAGTCCACCA	GACCTCAGTGTCTGCCCA	60	260
	8.1	CTTCTCCCTTGACCCTGGA	CCAGGCCTGGCAGGGGTG	60	448
	8.2	TAAGGCCCTCCCAACAGCA	TCTCCCCAGCAGAGACCAG	60	448
	9/10	ACATCCTTTAGACCCTCT	CCCAAGTCAGAAGGCTCA	55	464

Gene	Exon	Forward Sequence (5' → 3')	Reverse Sequence (5' → 3')	Annealing temperature (°C)	Product size (bp)
	11	CTCGACTGTTCTGTGTCC	CCTGTGGCTCACGGTTCA	60	377
	12	TGAACCGTGAGCCACAGG	TTGGCACAGCCTCTTCTC	55	305
	13	CAGAGCCCTGTGCTGAGT	CAGCCATTATGATACTGGC	55	284
	14/15	GCTCCGAACCAAGAGCCC	ATGGGGGTATGGCCTCTAC	60	379
	16	GACCTTAATTGCTAAGGG	TGCGCTGCAGCTGTGAGA	55	282
	17	TCAGAGTACAGCCTGCAGG	AGACAGCACAGCCACTCG	60	320
	18	CTCTCACGGGACTGTCA	AAAGGTGCCTCTGCCCAT	60	291
	19/20	CGACACAGCCATGTGGGC	GACCATGAGGAGGCAGTG	60	546
	21.1	GGGCGTTCAGGTAGACAG	CTCGTCCTCGTCCTCATC	55	582
	21.2	GAGGCCGACAGCACAAGTG	TCCTGGGTGCCGCTGGAAC	55	371
	22	GAGTCAGGGTTGCTTCTG	AGAAGCCAGGTTAGTGTC	55	217

Table 33 In house Sanger sequencing primer sequences

Primers were tagged with sequences specific for sequencing primers. The sequences for the forward and reverse primers were 'GTAGCGCGACGGCCAGT' and 'CAGGGCGCAGCGATGAC' respectively.

7.2.2. Diagnostic lab Sanger sequencing primer sequences

Gene	Exon	Forward Sequence (5' → 3')	Reverse Sequence (5' → 3')	Annealing temperature (°C)	Product size (bp)
CFH	1	GGAGTGCAGTGAGAATTGGG	CAACAATGTCAAAAGCCACTC	60	458
	2	CCTGTGACTGTCTAGGCATTTTT	TTTTGTATTTGACTGGCAATAGTGA	60	434
	3	CCTTACATTCAATCTGTCTTC	CCATCATAGTTAAACTTTCAGG	60	399
	4	GACACTCAGAATGGCATCGAG	GATCAGGCTGCATTCGTTTTTG	60	390
	5	CCTCCAATCTTATCCTGAGGATG	GCTGATATTCCTTAGAATGAACG	60	411
	6	CCTGATGGAAACAACATTTCTG	CATAAATTAGCACTCTACTTTTG	60	409
	7	GCCATTTTGTATTATGCTAAGG	CTTACTTTGTATATACAATAAGAC	60	442
	8	TTTCTTTCTGGAGAGCCTTCAT	TTTGGTCACTTTGCTTGAACA	60	658
	9	TTAGTAACTTTAGTTCGTCTTCAG	GGTCCATTGGTAAAACAAGGTG	60	486
	11	GGTTTTCAGTTACAAATGACTC	GAAATTATATCAGCCCCCAC	60	417
	12	GCTTTTTCTTCTTAGAATGGG	CAACATTCCTTAACAATTCCTC	60	395
	13	TCTGATGCCCCTCTGTATGACC	CTTATTTCAGCAATTGTAAGATAAG	60	358
	14	GAAGAAAATCTTTCCATTTTACTG	AAAATACAAAAGTTTTGACAAG	60	350
	15	ACTTGTCAAAACTTTTGTATTTTG	GGAAATGTTGAGGCATATCTG	60	405
	16	GATTAAGTCATAATTTTACCATGC	CACACATACCTATTACTTTTCC	60	392
	17	GTGATAATTTATGAAACAGTTATTGATCTTTC	GTGATTGATTAATGTGCCTAGGTC	60	424
	18	GTATTTTATTTGTTTTTAACCCTTTGA	CCTCACTTTGATAACAAGAGATTATTT	60	340
	19	GTGATGTCATAGTAGCTCCTG	CTAGAGTCCCTGTTTACTTTC	60	419
	20	CGCTATTTTAGAATCCATTACATG	GGCCTCCCAAAGTGCTGGG	60	379
	21	GTAACTGTTATCAGTTGATTTGC	TAACCCTGCTATACTCCCCC	60	332
	22	TTCTTCCAGGACTCATTTCTTTC	GTGAGTATTTTGTTACAAACAGTG	60	439
	23	CCTAATTCTCATACATTAAACATCG	GACACAACCGTTAGTTTTCCAGGA	60	653
CFI	1	GGAAACAAGTTCCTATTGGTC	CTTGTTCAGTTCAATGCCTCT	60	565
	2	AACACCATCCTCATCTGTACT	TCTGATAGAAAGTATGGCATAC	60	495
	3	CGTCATGATGTTCAAAGCTCAC	GGAAATTAATACACAGAAAGGTTAGG	60	395
	4	GTTTCTTACCTGCTACAGTAG	TATTATTGCCTCTGTGACTGG	60	403

	5/6	TCATGCCACCACTCCATAAA	TTTACCTAAAGAAAAGTTTCAGAATCC	60	525
	7	CAGCTAGACATCCTACAATTC	CATTAAAATATAAGACCAAAGAAC	60	307
	8	ACATGCCTTGGGGATTTTGTA	GCTTGAATCAATTATATATATGCAAATG	60	307
	9	TACTAATGATTCCAGCCTGTC	CTTATAATTACATCATCCTCCTGC	60	356
	10	GTCTTGCCATGGAAATATCTTAG	GCTTTATCATCTGCCACAATC	60	297
	11	AAAGTGAAGTGTCAGAATTCCTATTTA	GATGTTATGCTTCTCTCTGAGTGCT	60	553
	12	AAACACGCTAAGGAAGAGTTC	ATTAGAGGAAGAAACCTGAGC	60	464
	13	TTGTTAAATGCCATGGAGGAG	CGTTTTATTTCCATTAAATGGAACTC	60	403
MCP	1	CCGCCCTGGTGACTCGACGCACTT	GTCCCCGCATTCCCCGCCACATAC	60	493
	2	AATCTTGCCAAGGGCCTTTCTG	GTAGTGGAATATGTACCCCAAAATGTA	60	416
	3	TATATTCCCACCCATTCAAAAGA	GAAAGCAAAATTAAGTTTACAGGAAT	60	334
	4	GTTTAAGAAACCACCCCCTCAAAC	CTTAAAGTGTAAAGGGTGTAAAGGA	60	326
	5	AGAAACCCTATATTGACAAATTTATTG	AGGAAGCACATACACCTGCTTTG	60	433
	6	CATTCCTTGTCTCTGTTCACACTG	TTGTTTCTTTTCAGTTATTTTGTATGAC	60	428
	7/8	GTTCTTAGCACGTTATGTAC	ATGGCTATACAAATGTCCTC	60	583
	9	ACACCCATCCTCACATTACTTTCA	CTGGGGGAGGGATAGCATTAG	60	533
	10	CACAGAATGTTGTCACAGAAAATG	GGATCCTATGTTTGGGCACCTC	60	413
	11	AATTGCGATTCAAGATGAGATT	AAAAGCACATTTGGGAAGC	60	480
	12	ATTATGGGGAGTTGGATTTAGAT	TGTTTTTATGTTGCAGTGATGTTA	60	473
	13	GAACCTAATTCTCAGCTTTC	CCTCTTTGATATTTACTGCC	60	365
C3	1	GATAAAAAGCCAGCTCCAGC	CACCCTGAATTCTACAGGGAC	60	290
	2	CACATCCGTGGAATGACAA	GCTTAGAAAGGGAGAAGACAGA	60	373
	3/4	GGGACCCAGCCCAAGATC	TTCCGGTGTGTCTTTCTCTG	60	526
	5/6	GAGAATGTCCAGACACAGGTC	GGAGAGATGGCGTTGGT	60	489
	7	GGAGTCAGTGCCCAGAAC	TCCCTCACCTGGCTCTT	60	229
	8/9	GTCTTCAGCACAGGCAGGAG	ACCCCACTTTCACTCTGAGC	60	556
	10/11	TCTGAGGGAGGAGGTCTAAT	CTGTACCGTCTTCCCAGTG	60	514
	12	GATTCGGGTGTGGGTATCC	GGGATGGGGGAAGGAGTC	60	381
	13	GATCCCTCCCTGCGTTC	AGAGAGAGAGAGAGAGGAGTAG	60	365
	14	TCTCCCAGGGCTGACTT	GCCTCCGCCTCTTCTCA	60	279
	15	GGATCCCAATTGTCAGCCT	GTGGAGGTGCTGTCTGT	60	268

	16/17	GGAACACCCCCACCTCACAC	TCCCTCCTCAGACAGGAGT	60	610
	18/19	CATGAGCCATGGCAACTG	ATGACACTCAGACACCCT	60	604
	20/21	CCTGGCCTTGTCCACTC	AGTACGAAGACCAGGAGC	60	603
	22/23	TCTGTGTGTCTGTCTGTTCTT	GTTAAACACCTCCAGAATGAGAT	60	461
	24	AGTCCTCGCCTGTCCCTAAC	TCAAATGAGGGGAGTGGCTA	60	494
	25	TCCACCTCCTCGTTCTGATC	TGGACTCTGCAGGTCCAGG	60	423
	26	CATGGCAGTCTCTGGATCT	TCGTGTTCATCCTGCGAG	60	315
	27	CACCTAGAAGAGACTCAGCC	GACTGCAGTGATGTCTGTTATTG	60	379
	28	CAATGCCCACATGACCG	TTCAGCCATGCATCTCCC	60	276
	29	ATGCATCTCTTTCTGAGCTTTC	GGGCCTCAGTGTCTTCTCTA	60	366
	30/31	TCCCAGCTCTGATTTGAACC	CCACCTGGTAAGATGGGAGA	60	620
	32/33	GGAACAGAAACCCACACCTG	CTTGGAAAGTACTGAATATCATGGA	60	515
	34/35	TGCTGCTATGTGGGAATCAG	CCAGAGACAAAAGCTGAAAGG	60	534
	36	GAGGGATCTGAGGGGTTGAAG	TGAAGCCACACCATGACAAC	60	475
	37/38	GGAGGGAGGCCCTTATC	ACAATTGGACACACACAGC	60	507
	39/40	CCCTCATGGTCAACCTAGC	GGCGTGACAATGGTGTG	60	465
	41	CCACACCATTGTCACGC	CTGGGGATTTCAGCCTCT	60	339
CFB	1	GCAGGGGAAGGGAATGTGA	AAAGGCCAAGGAGGGAT	60	284
	2	TGCCTGATGCCCTTTATCT	CACCCTAAAACTGCTCCTAC	60	522
	3	TACACCTAAGGCAGCCTTTC	CGTCAGGGAGACAGCAA	60	319
	4	CTCTCTACCTTGCTCACGG	GGAGAGAAGACAGTAGGATGG	60	333
	5/6	CCTGACACTCCCAGACAT	GAAGATCAGCGAGATTCCATT	60	468
	7	GAAATCTCCCAATCACAGTATTCTAT	TCCACCCTGAACCTCCT	60	393
	8	GGTCAAAGGGAAGTCCGTG	GGAGAGATGAACAGCCAGCTA	60	439
	9/10	CCTTCCTCAACTTGCTCAC	GGTGTAGAGGAAGAATGAATTACT	60	519
	11/12	ATCCCGTGGTCTTTCCCTTTC	AGGATGGGTGGAGTGTAGGA	60	600
	13/14	TCCTACACTCCACCCATCCT	CCCCTAAGCCCTTCTAGAGC	60	630
	15/16	CTGGCCCAGAACCTAGCTC	CCAGAATCACCTGCAAGGAG	60	641
	17	GTTGTGCTACAAGTGCCCAAG	CATGCTGACCACTTGGCATC	60	194
	18	GCCATGCTTCCAGGATTAG	CCCAAAAGGATCTGGAACAC	60	328
THBD	1.1	CTGTGCCCCTCTGCTCC	TGGTGTTGTTGTCTCCCGTA	60	526

	1.2	TCATTTCCTTGCTACTGAACG	CACGCTGCAGTCCCAAG	60	582
	1.3	GCTCCCCTCGGCTTACAG	CGTCCACCAGGTCGTAGTTAG	60	413
	1.4	ATACTGGAGCCCAGTCCGTG	GTCACAGTCGGTGCCAATG	60	580
	1.5	GCACGGACATCGACGAG	TTTGGTAGCAAAGCTGGGG	60	576
DGKE	2.1	AGCGGAGCCACCTTCACT	GTCCAGGACCTTGGTGTCAT	60	522
	2.2	GTTCAGCCAGCCCACCTACT	GCAGGCCTCTACACCACTCT	60	374
	3	TGCCAAATGTTATGGTAAGTAGTGA	TTCCTGGCACCAATTCTTCT	60	420
	4	TCAAGGCATGGAAAATGTG	CAGGTGATTAGTGACTTGAAGCAT	60	310
	5	CAGATGAAGGAAAATGTTGGA	GGATGGCTTGAACCTGAGAA	60	430
	6	GCACAAGCTTTAGCAAAACAA	TGGCACATTTCATCTACACAAG	60	414
	7	TTTGCATGCTCATATACGTGTG	GTGCCGTATCTTGTGCAATG	60	272
	8	TGACTCAAGCTTAGGTGGAACC	AAAACAGAGCAGGTAAATATTTCTGTA	60	330
	9	GGGAGCTTTGATACAGCGTATTT	AGCATTTCTGTTTAAAAAGTCTTCA	60	275
	10	ATGAGAGGCCCTGAGTGAGA	GAAAATTTGAGGAAAAATGTTAAAGTC	60	401
	11	AAGTTGATGGTCCAACTGTGTT	GCTTAGAATTGGATTGGGACA	60	291
	12	GGTTAAGGGATTGTGGATGGT	GCTGGTTTGGGGATGCTAT	60	459

Table 34 Diagnostic Lab Sanger sequencing primers.

Primers were tagged with sequences specific for sequencing primers. The sequences for the forward and reverse primers were 'GTAGCGCGACGGCCAGT' and 'CAGGGCGCAGCGATGAC' respectively.

7.3. Appendix C

7.3.1. In house MLPA primer sequences

Gene	Exon	Left Primer (5' → 3')	Right Primer (5' → 3')
CFH	1	CACAATTCTTGGAAGAGGAGAACTG	GACGTTGTGAACAGAGTTAGCTGG
	2	TGACAGGTTCCTGGTCTGACCAA	ACATATCCAGAAGGCACCCAGGCT
	3	CGTTTTAGAAAGGCCCTGTGGACATCCTGGA	GATACTCCTTTTGGTACTTTTACCCTTACAGGAGG
	4	GGTATCAATTGCTAGGTGAGATTAATTACCGTGA	ATGTGACACAGATGGATGGACCAATGATATTCCTATATGTG AAGG TAG
	5	GTGCAATGGAACCAGATCGGGAATA	CCATTTTGGACAAGCAGTACGGTTTG
	6	CATGGGTTATGAATACAGTGAAAGAGGAGA	TGCTGTATGCACTGAATCTGGATGGCGTCCGTTGCCT
	7	GGTGACTACTCACCTTTAAGGATTAAACACAGAACTGGA GATGAAATCACGTAC	CAGTGTAGAAATGGTTTTTATCCTGCAACCCGGGGAAATAC AGC
	8	CACATTCATTGCACACAAGATGGA	TGGTCGCCAGCAGTACCATGCCTC
	9	GTACAGGGTAAATCTATAGAC	GTTGCCTGCCATCCTGGCTACG
	10	CAGAATGGGAAATGCTAATTCAGCTCCTCCAGGCAGCC CAATGGG	GCTGGTGGCTTTGAGATTATTAAACTCTTTCTCTGCTGC
	11	TGAAACATCAGGATCAATTACATGTGG	GAAAGATGGATGGTCAGCTCAACC
	12	GAATGCCAGAACTAAAAATGACTTCACATGGTTTAAG	CTGAATGACACATTGGACTATGAATGCCATGATGG
	13	ACCTAATTCCGTTCAGTGCTACC	ACTTTGGATTGTCTCCTGACCTCC
	14	CAGTGAAGTGGTGGAATATTATTGCAATC	CTAGATTTCTAATGAAGGGACC
	15	GGAGGAGAGTACCTGTGGAGATATACCTGAACTTGAAC ATGG	CTGGGCCCAGCTTTCTTCCCCTCCTTATTACTATGG
	16	GGAAAAGAAGGATGGATACACACAGTCTG	CATAAATGGAAGATGGGATCCAGAAGTGAACTGC
	17	GGCACAAATACAATTATGCCCACCTCCACCTCAGATTCC	CAATTCTCACAATATGACAACCACACTGAATTATCG

	18	CCATGTTCACAACCACCTCAGATAGAACAC	GGAACCATTAATTCATCCAGGTCTTCAC
	19	GGAATTGATGGGCCTGCAATTGCAAAATGCTTAG	GAGAAAAATGGTCTCACCCTCCATCATGC
	20	GGATGGAGCCAGTAATGTAACATG	CATTAATAGCAGATGGACAGGAAGGC
	21	CTTATATAGTGTCGAGACAGATGAGTAAATATCCATCTG GTGAG AGAGTACGTTATCAATGTAGGAGC	CCTTATGAAATGTTTGGGGATGAAGAAGTGATGTGTTTAAAT GGA AACTGGACGGAAC
	22	CGTAAGTACTTTAATATTCACGTGGCTG	GAAAAATCTCTGTGATGAGTCTGATATTTCACTGTTTG
	23	GGACAGCCAAACAGAAGCTTTATTC	GAGAACAGGTGAATCAGTTGAATTTGTG
CFHR5	1	GAGACTACCAAGCATGTTGCTCTTATTCAGTGTAATC	CTAATCTCATGGGTATCCACTGTTGGGGGAGAAGG
	2	CCCAGGAACACTTTGTGATTTTCCAAAAATACACCATG	GATTTCTGTATGATGAAGAAGATTATAACCC
	3	CAAATTATTTGCAACACAGGATACAGC	CTTCAAAACAATGAGAAAAACATTTCGTGTG
	4	GTCATGTTCCAATTTTAGAAG	CCAATGTAGATGCTCAGCC
	5	CCACCTCCTCAACTCTCCAATGGTGAAGTTAAG	GAGATAAGAAAAGAGGAATATGGACAC
	6	GGATACATACCTGAACTCGAGTACGGTTATGTTCAGCCG TCTGTC	CCTCCCTATCAACATGGAGTTTCAGTCGAGGTGAATTGC
	7	CAGATGTTCAGACATCTTCAGATACAG	GCACTCAGTCTGTATAAACGGG
	8	GGGAACAATTCTGCCCACCGCCACCTCAGATACCTAAT G	CTCAGAATATGACAACCACAGTGAATTATCAGGATGG
	9	CCCATTATCAGTATATCCTCCAGGGTCAACAGTGACGTA CCGTTGC	CAGTCCTTCTATAAACTCCAGGGCTCTGTAACTGTAACATG C
	10	GCTGTTGAATTCCAGTGTAAATTCC	CACATAAAGCGATGATATCATCACCACC

Table 35 In house MLPA probe hybridisation sequences for CFH and CFHR5.

7.4. Appendix D

7.4.1. Diagnostic Lab MLPA primer sequences

7.4.1.1. CFH, CFHR1, CFHR2, CFHR3 and CFHR5

Gene	Exon	Left Primer (5' → 3')	Right primer (5' → 3')
CFH	1	GAGTCTCTTTTAATCTTACCTTCTGCTACACAA	ATAGCCCATAACATAAGGCAAATAATCTTTGCTAGAAGTCTCATTTTTTGGATCT
	2	TCTGACAGGTTCCTGGTCTGACCA	AACATATCCAGAAGGCACCCAGGCTATCTATAAATGCCGCCCT
	3	TGGACATCCTGGAGATACTCCTTTTGGT	ACTTTTACCCTTACAGGAGGAAATGTGTTTGAATATGGTGTAAAAGCTGTG
	4	GGTATCAATTGCTAGGTGAGATTAATTACCGTGA	ATGTGACACAGATGGATGGACCAATGATATTCCTATATGTGAAGGTAGA
	6	CATGGGTTATGAATACAGTGAAAGAGGAGA	TGCTGTATGCACTGAATCTGGATGGCGTCCGTTGCCT
	9	CCTGTTATTTTCCTTATTTGGAAAATGGATATAATCAAAATC	ATGGAAGAAAGTTTGTACAGGGTAAATCTATAGACGTTGCCTGCCATCCTGG
	Intron 10	GAAAGAAGCTGGTGCAGATAGGTAGTCA	TATTTGGAACATCCAGAAAGTCAGTGACAAGTATGACTAACAGAAAATGCT
	Intron12	GCTTTGCAACTTTTGACTTGGACACATT	ATGATTGAGTCGCCAATGGAAAATCAAAACTGGATATGTGAAGTGTGGA
	13	CTGCAAACCAGGATTTACAATAGTTGGACCT	AATTCCGTTCAGTGCTACCACTTTGGATTGTCTCCTGACCTCCCA
	Intron15	CCCTCTACATCAGTGGTATAGCTGAGTGA	CATGAGGTAGTCAGGGACTGAGTCAGGACGTAAATCTCATTGAC
	18	CACCTCAGATAGAACACGGAACCATTA	ATTCATCCAGGTCTTCACAAGAAAGTTATGCACATGGGACTAAATTGAG
	23	TCAATACATAAATGCACCAAAAGTGATATCAATACATA	AATGCACCAAAACTGATGAAATGTAGATACTTCTACAAGATG
	Down-	GGAATAAGACAGACAAAACCACATGATTGCACTTATA	TGCAATCCGTTGTAAAATATGGTGACTTTAGTTAAC

	stream	CA	
CFHR1	Intron 1	CAGTAAAATATATTTCCCTCTGTTGAAGGATAATTCA	ATTGAAATGGGAATTCCTCCATTTTAATGACCTTTAAAATTTAAATGATCAAATCTAGG
	3	GCTGAACAATGCAAAATTCATCAGAGAGTTTCAG	GTCCATGTGTAGTGATAATAGACTATAGCAATGCTTCTCACACAGAAGTGT
	5	CATTCAGCAGAATGGTTGTTCAATAATCTGTGAT	TATTTTGTTACCAACAGCAAAATATCAGACTCATCACACTGATTTTTCCAG
	6	CCAAATTCAGCTGATTCACCTGTTCTCA	AATAAAGCTTCTGTTTGGCTGTCCACCTTAATGCTATGTTATAATTTTCCAT
	6	CCGAACTGGACACAACAGTTAGTTTTCCAAGT	TTTAATATGGTGCTTTTAAGAAGAGAGCCACCGGTCTCAGCTTA
CFHR2	Intron 1	CACCAAGGTCTAGGATACTAATTAAACTATGTCTGTACT	TGGAGTTTCGATCATTATGCTTTACCCTTTGATTTCCAAAAAGTT
	2	GAGAAGGATATGCCAGACAAGATCATAAA	CACTTGATAATCACAGGAGCAGTGACCAGAGGAGCTGG
	3	GAAGATAATCCCTTGAAGTTTAAGTAATACCTGTG	TGTGGTTTATAGTATCGGGTTAGTTGACAAGAAATGGTTACAAAACTGAT
	4	TCATTCCTGTTGTCAGTATATGCTCCAGG	TTCATCAGTTGAGTACCAGTGCCAGAACTTGTATCAACTTGAGGGTA
CFHR3	Up-stream	GTTTTCTTATGATTGCAGGATATTTAGTCCGAGGT	AGAAAGGGACATAAACTAAAGGAAATCATTTAAATC
	Up-stream	TTCATTTCTAAGTTCTAGGATATTCAGGGTGGTAATCT	TGGCTCTCAGTGGTTGCTTTTAAGTTTCTGTAGGAGGGAGAGAA
	1	TGGGTTTCCTGTGCTAATGGACAAGGTAAGTTA	AAAGAGATCTAAACACTCAGCTTCCCTCTTAAATGTAACTTCATGTAATATCTAG
	2	CGAGATAAATTTTGATATCACCTTCTCTCAAACATTTTCTTG	TGGAATTACAGCAACCATAACTAGTTTCCTGATCCTGACCTCT
	3	CATGTCTTTCTAAGTAACACGGACGACAG	TCTCAGACTTGTCTATATTAACTGTGGCAAAATGTTTTTGTCAACTTGTT
	Intron 4	CAATGTTATCATGAGTTTTGGGGGTTATATG	AATTCCTACATTTCTAGAATACTGTTTTCAATTTCTATACTTATCAAGGGCTCTGTG
	6	TAATAAGTTCAGTTGCACTTCCCCAACA	TCACAGCAGAGATCATCAGGTGAGGTGATAAACGTGAAC
	6	TCTGGGGTATTCCACTATCCCTTCCCG	ACACACTGCTTGAAATGATAGAATTGATGTATTTGCATTA

			TATCCCAA
CFHR5	1	CAGTGTAATCCTAATCTCATGGGTATCCA	CTGTTGGGGGAGAAGGTAAGTTGAAAACAGATCCGAATATTTTA
	2	CCAAAATACACCATGGATTTCTGTATGATGAAGAAGAT	TATAACCCTTTTTCCCAAGTTCCTACAGGGGAAGTTTTCTATTACTCCTGT
	3	GAAAAATGGTCATTCTGAATCTTCAGGACT	AATACATCTGGAAGGTGATACTGTACAAATTATTTGCAACACAGGATACAGC

Table 36 Diagnostic lab CFH, CFHR1, CFHR2, CFHR3 and CFHR5 MLPA probe hybridisation sequences.

SALSA MLPA probemix P236-A1 ARMD (MRC Holland).

7.4.1.2. MCP and CFI

Gene	Exon	Left Primer (5' → 3')	Right primer (5' → 3')
MCP	1	CCAATTCGGCTCTTGCCACGCCCACCTGTCCTGCA	GCACTGGATGCTTTGTGAGTTGGGGATTGTTGCGTCCCA
	Intron 1	GAGCAAAGGCACCAAGATGTGACACTA	CTGGACTGCCAAGCTATTCAGTGTGGCAGAGTAGGGC
	3	CCTGCAAATGGGACTTACGAGTTT	GGTTATCAGATGCACTTTATTTGTAATGAGGGGTAAGTTG CTCCTTAGA
	5	GTCGTGCTGCTCCAGAGTGTAAAGGTA	GTGTTTCAATTTATTTCCTTCTTCATTTGTAAATACTATGGA AACATTT
	6	GATATCTGTTTTCCATTTTCGACTACTGGA	AATCGACATTTGACCACTGGAAAATTAGAGAAACCTCAGA ATTAATGGA
	7	CTGCTGCCTCCATCTAGTACAAAACCTCCAGCT	TTGAGTCATTCAGGTTTAGTAGCTTCTTCCTTATATGTCTT CTTCCT
	8	TGAGGCACTGGACGCTGGAGATT	TTGTAGTGGAAGAAGTCGACACTGGAAAACAAAATAATAA TGTA
	9	CAGAGTGTGGTGAGCATTCTTGTACATA	CATCTTACATACTTGTGCAAATATATATGTAGGTTAAATTCT AAGATGTGG
	10	CGTGTAGGTAATTAGTTTTTATGAGGTGCCCA	AACATAGGATCCTTGGTAGGGTAAGATAACTTTCTTAAATG CTGTGT
	11	CTCAGATGTTTGGGTCATTGCTGTGAT	TGTTATTGCCATAGGTAAGTATCACAAATTTTGACACCACT TAAGT
	12	TGCAGTAATTTGTGTTGTCCCGTACAGATA	TCTTCAAAGGAGGAAGAAGAAAGGGTAAATTAAAGCATGT TTCT
	13	CTTCTCTGTGGGTCTCATCAGTTAGGT	ATGTGCTGAACAGAGGAAGAGAAAACGGCTGAATAAAATA GTGACTTCA
	14	CATAAGAAGTGAGAGGACTCTGACAGCCA	TAACAGGAGTGCCACTTCATGGTGCGAAGTGAACACTGTA GTCTTGT
CFI	1	TACGGTAAATCTCTCTGGATTTCAGCCAA	ATTCTTTCAGAGTTCAAAAGTACAAAGCTCTTTAGGAGGTT TGTTCT
	Intron 1	GAACCTACTCAGACAAGGAAACTAGGGGA	TTCAGAAAAGATGGAAAAGAAACCCATAGCACAGTGGGAG TGGCTGC
	2	GAATGGCACTGCAGTGTGTGCAACTA	ACAGGAGAAGCTTCCCAACATACTGTCAACAAAAGAGTTT

			GGAATGTCTT
	3	CATGCTCCAGCTGCTTTTGCATATGA	ACATTGTCTTATCTTGGTCCACAAGTTTTACTTCAACTATTCCCTCTGAATCT
	4	GTCTACATGTGCATTGCCGAGGAT	TAGAGACCAGTTTGGCTGAATGTACTTTTACTAAGAGAAGAACTATGGG
	5	GATGACTTCTTTCAGTGTGTGAATGGGA	AATACATTTCTCAGATGAAAGCCTGTGATGGTATCAATGATTGTGGAG
	6	CAAGGCAAAGGCTTCCATTGCA	AATCGGGTGTTTGCATTCCAAGCCAGTATCAATGCAATGGTG
	9	CAGTGCCTGGACATGTGGCA	TCCTTTATCAGCTAGAGGCTGGTAAATGTTTAACCTCTTGGCCAGCC
	10	TGGTGGCTGTTGGATTCTGACTGCT	GCACATTGTCTCAGGTAAAAGATGTCTTCAGTAAATTTTAGATTTCCC
	11	CCAGCCAGAAACGATGCATGTA	TCATTAGGTTGGAATAGGTAAGGAGACCAGGGGACACAGGCA
	12	CTGATGTGGTGGGAGGAGATGT	TTGATAGGGGAAATACATACATCTTGACATCTTGGATAAACCACTTGG
	13	CACCTGGGAACTCTGGTTTTCCACAGT	TTTCCCCCCAACTCACAACACCCCAGACATAAGTCACATT

Table 37 Diagnostic lab MCP and CFI MLPA probe hybridisation sequences.

SALSA MLPA probemix P296-A2 (MRC Holland).

7.5. Appendix E

7.5.1. WES raw data

7.5.1.1. Non-amplified samples

Family	ID	Affected	Mean coverage	Total number of reads received	Total number of reads mapped to reference genome (after removal of duplicates)	Percentage of exome covered				
						1x coverage	5x coverage	10x coverage	20x coverage	40x coverage
1	II:1	Y	77.5	114974982	98586935	97.3	95.9	94.3	89.1	68.3
	II:2	Y	85.4	134872396	114981021	97.1	95.9	94.6	91.0	75.4
2	II:2	Y	76.3	112217406	95704007	97.1	95.6	93.6	87.4	65.5
3	II:3	Y	70.4	124133766	104459515	97.0	95.5	93.7	87.3	63.9
	II:4	Y	76.8	130534982	109001220	97.1	95.7	94.0	88.6	68.3
4	I:1	Y	73.7	121353664	101385367	97.1	95.5	93.5	87.0	64.3
	II:1	Y	68.7	117968364	98693616	97.1	95.6	93.7	87.1	62.3
5	I:1	N	88.9	139361596	133479064	97.4	96.3	95.2	91.7	76.6
	I:2	N	82.6	121274658	116839474	97.0	95.8	94.6	90.7	73.5
	II:1	N	73.4	123569540	102987732	97.1	95.6	93.8	87.6	65.1
	II:2	Y	75.7	122146926	102046946	97.0	95.5	93.6	87.8	66.8
	II:3	Y	103.3	147386324	131463134	97.4	96.5	95.6	91.6	73.5
	II:4	N	101.5	142635116	126757457	97.0	96.2	95.2	90.8	70.7
6	II:1	Y	94.5	146038316	140311887	97.1	96.0	95.1	92.4	79.9
7	II:2	Y	87.1	134261126	128957859	97.4	96.3	95.3	91.8	75.4
	III:1	Y	82.0	124410752	119676674	97.0	95.8	94.7	90.9	73.5
8	II:2	Y	70.3	132421462	110974674	97.1	95.7	94.4	89.5	66.4
	III:1	N	78.2	136685376	113887522	96.9	95.7	94.4	89.9	70.2

9	II:1	Y	79.7	141398536	118269985	97.4	96.2	95.0	90.7	71.6
	II:2	Y	70.9	126935304	105904339	97.3	96.1	94.6	89.1	65.5
10	II:1	Y	61.6	66099336	64971496	99.0	96.2	90.4	73.3	39.5
	II:2	Y	94.1	126630412	114746149	96.9	96.0	94.9	90.2	68.2
	III:1	Y	93.1	127921188	116003397	96.9	96.0	94.8	89.7	66.6
	III:2	Y	86.4	119136342	108482050	97.3	96.3	94.8	87.9	61.7
11	VII:2	Y	100.7	139157236	125564757	97.4	96.5	95.3	90.4	69.9
12	II:2	Y	97.3	135221320	118392192	96.9	96.0	94.9	90.3	69.5
	III:1	Y	94.8	128595444	117158024	97.0	96.0	94.9	89.4	66.1
15	I:2	N	53.0	46400618	44141371	98.3	96.8	93.3	79.0	42.2
	II:1	Y	57.4	50080940	47759460	98.6	97.1	93.8	80.8	46.7
16	I:2	Y	57.9	63413530	62311388	99.0	95.9	89.5	70.9	36.2
	II:1	Y	67.6	70865504	69500893	99.1	96.7	91.7	76.9	44.4
17	II:1	Y	57.2	62877906	61710104	99.3	95.8	88.8	69.8	35.6
18	II:2	Y	63.7	70540038	69139986	99.1	96.4	90.8	74.5	41.1
	II:1	Y	60.2	68329296	67192618	99.4	97.0	91.5	74.1	38.9
19	II:1	Y	58.7	63293332	62185955	99.0	95.5	89.0	70.6	36.8
	II:2	Y	61.2	54735214	52129022	98.3	97.0	94.4	83.4	50.6
20	II:1	Y	63.9	70205374	68717548	99.2	96.5	91.0	75.2	42.1
	II:2	Y	54.8	62194670	61247621	99.3	96.1	89.1	69.5	34.5
21	II:2	Y	61.2	60207124	59064387	99.0	96.3	90.8	74.7	40.8
	II:1	Y	61.7	61956088	60824939	99.0	96.8	91.7	75.9	41.5
22	II:1	Y	58.8	58152316	57132907	99.4	96.9	91.4	74.5	39.3
23	II:1	Y	51.1	50865124	49916523	99.1	94.6	86.5	65.9	31.7
24	II:2	Y	78.5	78146778	76396557	99.2	97.6	94.3	83.8	55.0
	II:1	Y	56.4	57423640	56404657	98.9	95.8	89.5	71.3	36.3
25	I:2	Y	18.0	29944902	29976323	93.3	71.4	46.7	19.4	4.5
27	II:2	Y	73.1	70287754	67909660	98.4	95.2	89.7	78.3	53.5
Average			73.0	99723087.4	90507574.2	97.8	95.6	92.0	82.0	56.3
Standard Deviation			16.7	35933194.4	29725819.4	1.2	3.7	7.2	12.4	16.6

Table 38 WES raw data for non-amplified samples.

Y=Yes, N= No.

7.5.1.2. Whole Genome Amplified samples

Family	ID	Affected	Mean coverage (x)	Number of reads received	Number of reads mapped to reference genome (after removal of duplicates)	Percentage of exome covered				
						1x coverage	5x coverage	10x coverage	20x coverage	40x coverage
5	II:5	Y	35.4	42842532	42795642	47.8	23.4	18.6	14.3	10.2
11	VI:5	Y	64.4	66501440	65676781	98.6	94.4	87.3	69.1	39.4
13	II:7	Y	27.0	60670068	57010010	88.8	67.3	46.7	23.8	8.4
14	II:2	Y	42.1	82457622	75375855	94.8	87.1	74.1	48.3	20.5
Average			42.2	63117915.5	60214572	82.5	68.0	56.7	38.9	19.6
Standard Deviation			16.0	16355994	13824960.1	23.5	31.9	30.5	24.7	14.2

Table 39 WES raw data for Whole genome amplified samples

Y=Yes, N= No.

7.5.1.3. Sporadic patients

ID	Affected	Mean coverage (x)	Number of reads received	Number of reads mapped to reference genome (after removal of duplicates)	Percentage of exome covered				
					1x coverage	5x coverage	10x coverage	20x coverage	40x coverage
Sp4	Y	62.4	61141134	58873293.0	98.3	95.5	90.6	80.2	55.3
Sp6	Y	72.8	72516930	67704185.0	98.4	96.2	91.9	83.2	62.9
Average		67.6	66829032.0	63288739.0	98.4	95.8	91.2	81.7	59.1
Standard Deviation		7.3	8043902.5	6244383.6	0.0	0.5	1.0	2.1	5.3

Table 40 WES raw data for sporadic patients.

Y=Yes, N= No.

7.6. Appendix F

7.6.1. Mass Spectrometry coverage data.

7.6.1.1. 160kDa band

MRLLAKIICLMLWAICVA **EDCNELPPRRNTEILTGSWSDQTYPEGTQAIYKCRPGYRSLG**
LEADER SEQUENCE **FH CCP1**
NV IMVCRKGEWVALNPLRKCQKRPCGHPGDTPFGTFTLTGGNVFEYGVKAVYTCNEGYQL
FH CCP2
LGE **INYRECDTDGWTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCNS**
FH CCP3
GYKIEGD EEMHCSDDGFWSKEKPKCVEISCKSPDVINGSPISQKIIYKENERFQYKCNMG
FH CCP4
YEYS **ERGDAVCTESGWRPLPSCEEKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRNGFYP**
FH CCP5
AT RGNTAKCTSTGWIPAPRCTLKPCDYPDIKHGGLYHENMRRPYFPVAVGKYYSYYCDEH
FH CCP6
FETPSGS **YWDHIHCTQDGWSPAVPCLRKCYFPYLENGYNQN**Y**GRKFVQGKSIDVACHPGY**
H402Y **FH CCP7**
ALPKA QTTVTCMENGWSPTPRCIRVKTCSKSSIDIENGFISESQYTYALKEKAKYQCKLG
FH CCP8
YVTADGE **TSGSITCGKDGWSAQPTCIKSCDIPVFMNARTKNDFTWFKLNDTLDYECHDGY**
FH CCP9
ESNTGST TGSIVCGYNGWSDLPICYERECELPKIDVHLVPDRKKDQYKVGEVLKFSCKPG
FH CCP10
FTIVGP **NSVQCYHFGLSPDLPICKEQVQSCGPPPELLNGNVKEKTKEEYGHSEVVEYYCN**
FH CCP11
PRFLMKGP NKIQCVDGEWTTLPVCIVEESTCGDIPELEHGWAQLSSPPYYYGDSVEFNCS
FH CCP12
ESFTMIGHRSI **TCIHGVWTQLPQCVAIDKLKKCKSSNLIILEEHLKNKKEFDHNSNIRYR**
FH CCP13
CRGKEGWIHT VCINGRWDPEVNCSMAQIQLCPPPPQIPNSHNMTTTLNYRDGEKVSVLCQ
FH CCP14
ENYLIQE **GEEITCKDGRWQSIPLCVEKIPCSQPPQIEHGTINSSRSSQESYAHGTKLSYT**
FH CCP15
CEGGFRIS EENETTCYMGKWSSPPQCEGLPCKSPPEISHGVVAHMSDSYQYGEEVTYKCF
FH CCP16
EGFGIDG **PAIAKCLGEKWSHPPSCIKTDCLSLPSFENAIPMGEKKDVYKAGEQVTYTCAT**
FH CCP17
YYKMDGASNVTCINSRWTGRPTCR/QVKPCDFPDIKHGGLFHENMRRPYFPVAVGKYYSYY
FHR3 CCP1
CDEHFETPSGSYWDYIHCTQNGWSPAVPC **LRKCYFPYLENGYNQNYGRKFVQGNSTEVAC**
FHR3 CCP2
HPGYGLPKAQTTVTCTEKGWSPTPRCI RVRTCSKSDIEIENGFISESSSIYILNKEIQYK
FHR3 CCP3
CKPGYATADGNSSGSITCLQNGWSAQPIC **INSSEKCGPPPPISNGDTTSFLLKVYVPQSR**
FHR3 CCP4
VEYQCQPYYELQGSNYVTCSNGEWSEPPRCI HPCIITEENMNKNNIKLKGRSDRKYYAKT
FHR3 CCP5
GDTIEFMCKLGYNANTSILSFQAVCREGIVEYPRCE

7.6.1.2. 150kDa band

MRLLAKIICLMLWAICVA **EDCNELPPRRNTEILTGSWSDQTYPEGTQAIYKCRPGYRSLG**
LEADER SEQUENCE **FH CCP1**
NV IMVCRKGEWVALNPLRKCQKRPCGHPGDTPFGTFTLTGGNVFEYGVKAVYTCNEGYQL
FH CCP2
LGE **INYRECDTDGWTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCNS**
FH CCP3
GYKIEGD EEMHCSDDGFWSKEKPKCVEISCKSPDVINGSPISQKIIYKENERFQYKCNMG
FH CCP4
YEYS **ERGDAVCTESGWRPLPSCEEKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRNGFYP**
FH CCP5
AT RGNTAKCTSTGWIPAPRCTLKPCDYPDIKHGGLYHENMRRPYFPVAVGKYYSYYCDEH
FH CCP6
FETPSGS **YWDHIHCTQDGWSPAVPCLRKCYFPYLENGYNQNYGRKFVQGKSIDVACHPGY**
Y402H **FH CCP7**
ALPKA QTTVTCMENGWSPTPRCIRVKTCSKSSIDIENGFISESQYTYALKEKAKYQCKLG
FH CCP8
YVTADGE **TSGSITCGKDGWSAQPTCIKSCDIPVFMNARTKNDFTWFKLNDTLDYECHDGY**
FH CCP9
ESNTGST TGSIVCGYNGWSDLPICYERECELPKIDVHLVPDRKKDQYKVGEVLKFSCKPG
FH CCP10
FTIVGP **NSVQCYHFGLSPDLPICKEQVQSCGPPPELLNGNVKEKTKEEYGHSEVVEYYCN**
FH CCP11
PRFLMKGP NKIQCVDGEWTTLPVCIVEESTCGDIPELEHGWAQLSSPPYYYGDSVEFNCS
FH CCP12
ESFTMIGHRSI **TCIHGVWTQLPQCVAIDKLKKCKSSNLIILEEHLKNKKEFDHNSNIRYR**
FH CCP13
CRGKEGWIHT VCINGRWDPEVNCSMAQIQLCPPPPQIPNSHNMTTTLNYRDGEKVSVLCQ
FH CCP14
ENYLIQE **GEEITCKDGRWQSIPLCVEKIPCSQPPQIEHGTINSSRSSQESYAHGTKLSYT**
FH CCP15
CEGGFRIS EENETTCYMGKWSSPPQCEGLPCKSPPEISHGVVAHMSDSYQYGEEVTYKCF
FH CCP16
EGFGIDG **PAIAKCLGEKWSHPPSCIKTDCLSLPSFENAIPMGEKKDVYKAGEQVTYTCAT**
FH CCP17
YYKMDG ASNVTCINSRWTGRPTCRDTSCVNPPTVQNAYIVSRQMSKYPSGERVRYQCRSP
FH CCP18
YEMFG DEEVMCLNGNWTEPPQCKDSTGKCGPPPPIDNGDITSFPLSVYAPASSVEYQCQN
FH CCP19
LYQLEG NKRITCRNGQWSEPPKCLHPCVISREIMENYNIALRWTAKQKLYSRTGESVEFV
FH CCP20
CKRGYRLSSRSHTLRTTCWDGKLEYPTCAKR

7.6.1.3. 120kDa Band

MRLLAKIICLMLWAICVA **EDCNELPPRRNTEILTGSWSDQTYPEGTQAIYKCRPGYRSLG**
LEADER SEQUENCE **FH CCP1**
NV IMVCRKGEWVALNPLRKCQKRPCGHPGDTPFGTFTLTGGNVFEYGVKAVYTCNEGYQL
FH CCP2
LGE **INYRECDTDGWTNDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCNS**
FH CCP3
GYKIEGD EEMHCSDDGFWSKEKPKCVEISCKSPDVINGSPISQKIIYKENERFQYKCNMG
FH CCP4

YEYS **ERGDAVCTESGWRPLPSCEEKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRNGFYP**
FH CCP5
AT RGNTAKCTSTGWIPAPRCTLKPCDYPDIKHGGLYHENMRRPYFPVAVGKYYSYYCDEH
FH CCP6
FETPSGS **YWDHIHCTQDGWSPAVPCLRKCYFPYLENGYNQNYGRKFVQGKSIDVACHPGY**
H402Y **FH CCP7**
ALPKA QTTVTCMENGWSPTPRCIRVKTCSKSSIDIENGFISESQYTYALKEKAKYQCKLG
FH CCP8
YVTADGE **TSGSITCGKDGWSAQPTCIKSCDIPVFMNARTKNDFTWFKLNDTLDYECHDGY**
FH CCP9
ESNTGST TGSIVCGYNGWSDLPICYERECELPKIDVHLVPDRKKDQYKVGEVLKFSCKPG
FH CCP10
FTIVGP **NSVQCYHFGLSPDLPICKEQVQSCGPPPELLNGNVKEKTKEEYGHSEVVEYYCN**
FH CCP11
PRFLMKGP NKIQCVDGEWTTLPVCIVEESTCGDIPELEHGWAQLSSPPYYYGDSVEFNCS
FH CCP12
ESFTMIGHRSI **TCIHGVWTQLPQCVAIDKLKKCKSSNLIILEEHLKNKKEFDHNSNIRYR**
FH CCP13
CRGKEGWIHT VCINGRWDPEVNCSMAQIQLCPPPPQIPNSHNMTTTLNYRDGEKVSVLCQ
FH CCP14
ENYLIQE **GEEITCKDGRWQSIPLCVEKIPCSQPPQIEHGTINSSRSSQESYAHGTKLSYT**
FH CCP15
CEGGFRIS EENETTCYMGKWSSPPQCEGLPCKSPPEISHGVVAHMSDSYQYGEEVTYKCF
FH CCP16
EGFGIDG **PAIAKCLGEKWSHPPSCIKTDCLSLPSFENAIPMGEKKDVYKAGEQVTYTCAT**
FH CCP17
YYKMDGASNVTCINSRWTGRPTCR/QVKPCDFPDIKHGGLFHENMRRPYFPVAVGKYYSYY
FHR3 CCP1
CDEHFETPSGSYWDYIHCTQNGWSPAVPC **LRKCYFPYLENGYNQNYGRKFVQGNSTEVAC**
FHR3 CCP2
HPGYGLPKAQTTVTCTEKGWSPTPRCI RVRTCSKSDIEIENGFISESSSIYILNKEIQYK
FHR3 CCP3
CKPGYATADGNSSGSITCLQNGWSAQPIC **INSSEKCGPPPPISNGDTTSFLLKVYVPQSR**
FHR3 CCP4
VEYQCQPYYELQGSNYVTCSNGEWSEPPRCI HPCIITEENMNKNNIKLKGRSDRKYYAKT
FHR3 CCP5
GDTIEFMCKLGYNANTSILSFQAVCREGIVEYPRCE

Highlighted = Identified by mass spectrometry. ‘/’ = Break point

7.7. Appendix G

7.7.1. Patient clinical data

Family/ Sp Number	ID	C3 (0.68-1.38 g/L)	C4 (0.18-0.6g/L)	FH (0.35-0.59 g/L)	FI (38-58mg/L)	MCP FACS	Anti-FH antibodies	Genes screened	Other tests
1	II:1	0.63	0.29	0.44	68	-	Negative	C3, MCP, CFH, CFB, CFI, THBD.	AP100-N CH100-N CP100-N
	II:2	-	-	-	-	-	Negative	THBD	-
2	II:2	0.99	0.22	0.55/0.53	49		Negative	CFI, CFB, CFH18-23	-
3	II:3	-	-	-	-	-	-	CFH 18-23	-
	II:4	-	-	-	-	-	-	CFH 18-23	-
4	I:1	-	-	0.67	47	-	Negative		-
	II:1	-	-	0.61	46	-	Negative	CFH	-
5#	I:1	-	-	0.55	53	-	-	-	AP100-N CH100-N CP100-N
	I:2	0.67	-	-	56	-	-	-	AP100-N CH100-N CP100-N
	II:1	0.65	-	-	52	-	-	-	AP100-N CH100-N CP100-N
	II:2	-	-	-	-	-	-	CFH, CFI	-
	II:3	-	-	-	-	-	-	CFH18-23	-
	II:4	-	-	0.47	46	-	-	-	AP100-N CH100-N CP100-N
	II:5	-	-	-	-	-	-	-	-
6	II:1	1.66	0.29	0.97	77	-	-	CFH18-23, MCP	-

Family/ Sp Number	ID	C3 (0.68-1.38 g/L)	C4 (0.18-0.6g/L)	FH (0.35-0.59 g/L)	FI (38-58mg/L)	MCP FACS	Anti-FH antibodies	Genes screened	Other tests
7	II:2	-	-	-	-	-	-	MCP, CFH, CFI, CFB	-
	III:1	-	-	-	-	-	Negative	-	-
8	II:2	1.01	0.14	0.55	57		-	MCP, CFH, CFI	-
	III:1	1.41	0.12	0.76	68	Normal	Negative	C3, CFB, CFH18-23	-
9	II:2	1.18	0.10	0.72	58	-	Negative	MCP, CFH1-17, CFI	-
	II:1	0.63	-	-	53	-	Negative	CFH	-
10	II:1	1.63	0.55	0.77	83	-	Negative	-	-
	II:2	0.98	0.28	0.58	39	-	Negative	MCP, CFH18-23	-
	III:1	1.79	0.4	0.74	61	-	Negative	-	-
	III:2	1.71	0.53	0.87	70	-	Negative	-	-
11#	VII:2	1.14	0.16	0.56	49	-	-	MCP, CFB, CFH, CFI	-
	VI:5	-	-	-	-	-	-	CFH18-23	-
12	II:2	1.73	0.32	0.68/0.83	85	Normal	Negative	CFI	-
	III:1	1.23	0.21	0.13/0.57	56	Normal	++	CFI	-
13#	II:7	-	-	-	-	-	-	MCP, CFH, CFI, DGKE	-
14	II:2	-	-	-	-	-	-	CFI, CFH18-23	-
15#	I:2	-	-	-	-	-	-	-	-
	II:1	-	-	-	-	-	-	-	-
16	I:2	0.66	0.29	0.47	61	Normal	Negative	C5 R885H, C3, MCP, CFH, CFI, CFB, DGKE	-
	II:1	0.56	0.19	0.39	47	-	Negative	C5 R885H, C3, MCP, CFH, CFI, CFB, DGKE	-
17	II:1	-	-	-	-	-	Negative		-
18	II:2	-	-	0.70/0.68	63	-	Negative	CFH18-23	-
	II:1	-	-	0.70	-	-	-	CFH18-19	-
19	II:1	-	-	0.7	-	-	Negative	-	-
	II:2	-	-	0.6	-	-	Negative	CFH18-23	-
20	II:1	-	-	-	-	-	-	-	-
	II:2	1.25	0.33	0.57	51	Normal	-	-	-

Family/ Sp Number	ID	C3 (0.68-1.38 g/L)	C4 (0.18-0.6g/L)	FH (0.35-0.59 g/L)	FI (38-58mg/L)	MCP FACS	Anti-FH antibodies	Genes screened	Other tests
21	II:2	1.34	0.44	1.03	72	-	-	-	-
	II:1	1.2	0.51	1.05	60	-	-	CFH18-23	-
22	II:1	0.75	0.20	0.50	49	-	-	C3, CFH, CFI	-
23	II:1	-	-	-	-	-	-	CFH	-
24	II:2	-	-	0.61	-	-	Negative	CFH18-23	-
	II:1	-	-	-	-	-	Negative	CFH,CFI	-
25	I:2	-	-	0.56	-	-	Negative	CFI, CFB	-
26#	II:6	0.7	-	-	-	-	-	-	-
27	II:2	1.33	-	0.45	-	-	-	CFI	-
28	II:2	-	0.28	0.32	-	-	Negative	-	-
Sp1#	II:1	1.01	0.22	-	-	-	-	CFH	-
Sp2#	II:1	1.64	0.81	0.72	69	Normal	-	CFH, CFI	-
Sp3#	II:1	-	-	-	-	-	-	-	-
Sp4	II:1	0.97	0.26	0.50	60	-	Negative	C3, MCP, CFH, CFI, CFB, DGKE	-
Sp5	II:1	1.24	0.18	0.71	81	-	-	-	-
Sp6	II:1	0.67	0.19	0.46	62	Normal	-	C3, MCP, CFH, CFI, CFB, DGKE	-

Table 41 Clinical results for all patients.

#= consanguineous. '-'= not available.

7.8. Appendix H

7.8.1. Patient screening data

Family/Sp Number	ID	Exome preparation	In house MLPA		NHS MLPA			Western blot		
			CFH	CFHR5	CFHR1/3	CFI	MCP	CFH	CFHR1/2, FHL	CFHR4/5
1	II:1	1	N	-	2	N	N	N	N	N
	II:2	1	N	-	2	-	-	N	N	-
2	II:2	1	N	N	1	-	-	N	-	N
3	II:3	1	-	-	-	N	N	-	-	-
	II:4	-	-	-	-	-	-	-	-	-
4	I:1	1	-	-	2	-	-	-	-	-
	II:1	1	-	-	2	-	-	N	-	N
5#	I:1	1	N	N	-	-	-	N	-	-
	I:2	1	N	N	-	-	-	N	-	-
	II:1	1	-	-	-	-	-	N	-	N
	II:2	1	-	N	-	-	-	-	-	-
	II:3	1	-	N	2	N	N	-	-	-
	II:4	1	-	-	-	-	-	-	-	-
	II:5	2***	-	-	-	-	-	-	-	-
6	II:1	1	N	-	2	N	N	N	-	-
7	II:2	1	N	N	-	-	-	-	-	-
	III:1	1	-	-	1	-	-	N	-	N
8	II:2	1	N	N	-	N	N	N	-	N
	III:1	1	N	-	0	-	-	-	-	-
9	II:2	1	-	N	0	N	N	N	-	N
	II:1	1	-	-	1	-	-	N	-	N

Family/Sp Number	ID	Exome preparation	In house MLPA		NHS MLPA			Western blot		
			CFH	CFHR5	CFHR1/3	CFI	MCP	CFH	CFHR1/2, FHL	CFHR4/5
10	II:1	2	-	N	1	-	-	-	-	-
	II:2	1	-	N	1	N	N	N	-	N
	III:1	1	N	N	0	-	-	N	-	N
	III:2	1	N	N	-	-	-	N	-	N
11#	VI:5	2**	-	-	2	-	-	-	-	-
	VII:2	1	N	-	-	N	N	N	-	-
12	II:2	1	N	N	1	-	-	N	-	N
	III:1	1	N	-	0	N	N	-	-	-
13#	II:7	1*	-	-	-	-	-	-	-	-
14	II:2	1*	-	-	-	N	N	-	-	-
15#	I:2	3	N	-	-	-	-	-	-	-
	II:1	3	-	N	-	N	N	-	-	-
16	I:2	2	-	-	-	N	N	-	-	-
	II:1	2	-	-	-	N	N	-	-	-
17	II:1	2	-	-	-	N	N	-	-	-
18	II:1	2	-	-	2	-	-	-	-	-
	II:2	2	-	-	2	-	-	N	-	-
19	II:1	2	-	-	-	-	-	N	-	-
	II:2	3	-	-	-	N	N	N	-	-
20	II:1	2	-	-	-	N	N	-	-	-
	II:2	2	-	-	-	-	-	-	-	-
21	II:2	2	-	-	-	-	-	-	-	-
	II:1	2	-	-	-	N	N	-	-	-
22	II:1	2	-	-	1	-	-	N	-	-
23	II:1	2	-	-	1	-	-	-	-	-
24	II:2	2	-	-	1	-	-	-	-	-
	II:1	2	-	-	1	-	-	-	-	-

Family/Sp Number	ID	Exome preparation	In house MLPA		NHS MLPA			Western blot		
			CFH	CFHR5	CFHR1/3	CFI	MCP	CFH	CFHR1/2, FHL	CFHR4/5
25	I:2	2	-	-	1	-	-	-	-	-
26#	II:6	-	-	-	-	-	-	-	-	-
27	II:2	3	-	-	-	-	-	-	-	-
28	II:2	-	-	-	-	-	-	-	-	-
Sp1#	II:1	-	-	-	-	-	-	-	-	-
Sp2#	II:1	-	-	-	-	-	-	-	-	-
Sp3#	II:1	-	-	-	-	-	-	-	-	-
Sp4	II:1	3	-	-	-	-	-	-	-	-
Sp5	II:1	-	-	-	-	-	-	-	-	-
Sp6	II:1	3	-	-	-	-	-	-	-	-

Table 42 Screening results for all patients.

*1= Illumina TruSeq chemistry, 2= Illumina Nextera rapid capture exome, 3= Agilent SureSelect^{XT} Human all exon V5, *= WGA Sigma, **= WGA Qiagen repli-g, ***= WGA Qiagen repli-g FFPE, #= Consanguineous, N= normal.*

7.9. Appendix I

7.9.1. In silico data

Gene	Amino acid change	Family or Sporadic number	*In silico* testing							
			PolyPhen-2		Mutation Taster	Mutation Assessor	FATHMM	Radial SVM	GERP++	PhyloP
			HDIV	HVAR						
CFH	E625*	23	NA	NA	D	NA	NA	NA	5.12	2.876
	C1152S	28	D	D	P	H	D	NA	NA	2.737
MCP	Y189D	4	D	D	N	H	T	D	4.85	2.158
	c.286+2T>G	18	NA	NA	D	NA	NA	NA	2.33	0.661
	c.286+2T>G	22	NA	NA	D	NA	NA	NA	2.33	0.661
	F246I	26	D	D	P	M	T	NA	NA	0.729
CFI	I416L	27	D	B	D	M	D	D	5.62	2.145
ADAMTS13	R7W	7	B	B	P	N	D	T	-6.55	-1.237
	D217H		D	D	D	L	D	D	4.69	2.161
	I673F		D	D	D	M	T	T	1.6	0.027
	A1033T		D	D	D	L	T	T	5.59	2.64
	R1060W		D	D	D	M	T	T	3.17	1.192
	Q448E	24	B	B	P	N	T	T	4.25	0.678
	R507Q		D	D	D	M	T	T	5.42	2.535
	A900V		B	B	P	N	T	T	4.55	0.971
DGKE	T533P	20	D	D	D	L	T	T	5.8	2.216
	R155G	8	B	D	D	M	D	D	4.56	0.936
	L476P		D	D	D	M	T	T	5.36	2.15
	M1L	Sp1	B	B	D	NA	T	NA	NA	3.313
	W322*	Sp2	NA	NA	D	NA	NA	NA	NA	5.709
	K109E	Sp3	D	D	D	M	D	NA	NA	4.553
INF2	V102D	16	D	D	D	M	D	D	4.76	1.777
INF2	R950W	9	D	D	D	N	T	T	2.53	0.91
	R177H	Sp4	D	D	D	M	D	D	4.48	2.027

Gene	Amino acid change	Family or Sporadic number	*In silico* testing							
			PolyPhen-2		Mutation Taster	Mutation Assessor	FATHMM	Radial SVM	GERP++	PhyloP
			HDIV	HVAR						
C9	P167S	1,Sp5	D	D	D	M	T	T	3.81	1.401
	G126R	Sp6	D	D	D	M	D	D	5.32	2.484

Table 43 In silico predictions for sequence variants.

Table shows the in silico predictions for the sequence variants reported in this project. Predictions could not be carried out on all variants, therefore are labelled NA (not available). Software is not able to calculate predictions for INDELs, so they were not included. B= Benign, D= deleterious, H= high, L= low, M= medium, N= neutral, P= polymorphism and T= tolerated.

7.10. Appendix J

7.10.1. Known-TMA gene list

CFH
CFI
MCP
CFB
C3
THBD
ADAMTS13
MMACHC

Table 44 List of genes known to cause TMA.

7.10.2. CMT- associated genes

AARS	KARS
AIF	KIF1B
AIFM1	KIF5A
ALS11	LITAF
AMOXAD	LMNA
AMY1	LRSAM1
APOA2	MARS
AT3	MED25
BSCL2	MFN2
C12orf65	MPZ
CADM4	MTMR2/13
CHM	NDRG1
DHTKD1	NEFL
DNM2	PDK3
DYNC1H1	PLEKHG5
EGR2	PMP22
FBLN5	PRPS1
FGD4	PRX
FIG4	RAB7
GARS	SBF1/2
GBA	SCN4A
GDAP1	SH3TC2
GJB1	SIMPLE
HADHB	SURF1
HARS	SMAD1
HSP27	TRIM2
HSPB1	TRPV4
HSPB8	TRPVR
INF2	YARS

Table 45 List of genes associated with CMT.

7.10.3. Genes associated with a renal phenotype

ACTN4	LAMB2
ADCK4	LMX1B

ADD3	MAL
ALG1/13	MPV17
ALMS1	MPZ
ANLN	MTTL1
APOE	MTTY
APOL1	MYH9
ARHGAP24	MYO1E
ARHGDIA	NPHS1/2
CBS	NEU1
CD151	NXF5
CD2AP	PAX2
CDC42	PDSS2
CLCN5	PDCN
COL4A3/4/5	PLAUR
COQ2/6	PLCE1
CRB2	PODXL
CUBN	PMM2
CYP11B2	PTPRO
DCLRE1B/C	RAG1/2
DGKE	SCARB2
E2F3	SEPT7
EHD3	SHROOM3
EMP2	SLC17A5
FN1	SLC35A2
FOXC2	SMAD7
GLA	SMARCAL1
INF2	SYNPO
ILK	TCF21
ITGA3	TRPC6
ITGB4	TREX1
JAG1	TTC21B
JAK2	UMOD
KAT2B	WT1
KLF6	ZMPSTE24

Table 46 List of genes associated with a renal phenotype.

7.10.4. Complement gene list

C1QA	C9
C1QB	CD55
C1QBP	CD59
C1QC	CFD
C1R	CFP
C1S	CLU
C2	CPB2
C4A	CR1
C4B	CR2
C4BPA	FCN1/2/3
C4BPB	MASP1/2
C5	MBL2
C5AR1	SERPING1
C6	SFTPA1/2

C7	SFTPB
C8A	SFTPC
C8B	SFTPD
C8G	

Table 47 List of complement genes

7.10.5. Coagulation gene list

CPB2	KNG1
F10	PF4
F11	PLAT
F12	PLAU
F13A1	PLAUR
F13B	PLG
F2	PLGRKT
F2R	PROC
F3	PROS1
F5	PROZ
F7	SERPINB2
F8	SERPINC1
F9	SERPINE1
FGA	SERPINE2
FGB	SERPINF1
FGG	SERPINF2
HRG	TFPI
KLK1	VWF
KLKB1	

Table 48 List of coagulation genes.

7.11. Appendix J

7.11.1. Published hybrid CFH/CFHR3 paper

BRIEF COMMUNICATION www.jasn.org

A *De Novo* Deletion in the Regulators of Complement Activation Cluster Producing a Hybrid Complement Factor H/Complement Factor H–Related 3 Gene in Atypical Hemolytic Uremic Syndrome

Rachel C. Challis,* Geisilaine S.R. Araujo,* Edwin K.S. Wong,* Holly E. Anderson,* Atif Awan,† Anthony M. Dorman,‡ Mary Waldron,† Valerie Wilson,* Vicky Brocklebank,* Lisa Strain,* B. Paul Morgan,§ Claire L. Harris,§ Kevin J. Marchbank,‖ Timothy H.J. Goodship,* and David Kavanagh*

*Institutes of Genetic Medicine and ‖Cellular Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom; †Department of Nephrology, Our Lady's Children's Hospital, Crumlin, Dublin; ‡Department of Renal Pathology, Beaumont Hospital, Royal College of Surgeons in Ireland, Dublin, Ireland; and §Institute of Infection and Immunity, Cardiff University School of Medicine, Cardiff, United Kingdom

ABSTRACT

The regulators of complement activation cluster at chromosome 1q32 contains the complement factor H (*CFH*) and five complement factor H–related (*CFHR*) genes. This area of the genome arose from several large genomic duplications, and these low-copy repeats can cause genome instability in this region. Genomic disorders affecting these genes have been described in atypical hemolytic uremic syndrome, arising commonly through nonallelic homologous recombination. We describe a novel *CFH/CFHR3* hybrid gene secondary to a *de novo* 6.3-kb deletion that arose through microhomology–mediated end joining rather than nonallelic homologous recombination. We confirmed a transcript from this hybrid gene and showed a secreted protein product that lacks the recognition domain of factor H and exhibits impaired cell surface complement regulation. The fact that the formation of this hybrid gene arose as a *de novo* event suggests that this cluster is a dynamic area of the genome in which additional genomic disorders may arise.

J Am Soc Nephrol 27: ●●●–●●●, 2015. doi: 10.1681/ASN.2015010100

The complement factor H (*CFH*) and complement factor H–related (*CFHR1–CFHR5*) genes reside in a 360-kb region in the regulators of complement activation (RCA) cluster on chromosome 1q32 (Figure 1A).[1] This is an area of the genome that arose from several large genomic duplications,[2,3] and these low-copy repeats can cause genome instability in this region.

Mutations in *CFH* are the most common genetic predispositions to the thrombotic microangiopathy: atypical hemolytic uremic syndrome (aHUS).[4] The majority of mutations in *CFH* occur at the C-terminal end responsible for host polyanion binding.[5–7] Several of these *CFH* mutations (*e.g.*, S1191L and V1197A) have been shown to be gene conversion events between *CFH* and *CFHR1*.[8] Nonallelic homologous recombination involving *CFH* and *CFHR1* can result in *CFH/CFHR1* hybrid genes (Supplemental Figure 1).[9,10] A single family with a hybrid *CFH/CFHR3* gene (factor H [FH] protein complement control protein modules [CCPs] 1–19 and factor H–related 3 [FHR3] CCP1–CCP5) has been reported to arise through microhomology–mediated end joining (MMEJ)[11] (Supplemental Figure 1). These genetic variants have all been shown to impair C3b/host polyanion binding on host cells, thus impairing local complement regulation.[5,8,11]

Recently, a reverse *CFHR1/CFH* hybrid gene arising through nonallelic homologous recombination has been described in aHUS. This encodes an FHR1, where the C-terminal CCPs are replaced by the C-terminal CCPs of FH.[12,13] Unlike previously reported hybrid proteins, this does not impair FH cell surface binding but instead, acts as a competitive inhibitor of FH.[13]

Received January 27, 2015. Accepted September 7, 2015.

R.C.C., G.S.R.A., and E.K.S.W. contributed equally to this work.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. David Kavanagh, Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK. Email: david.kavanagh@ncl.ac.uk

Copyright © 2015 by the American Society of Nephrology

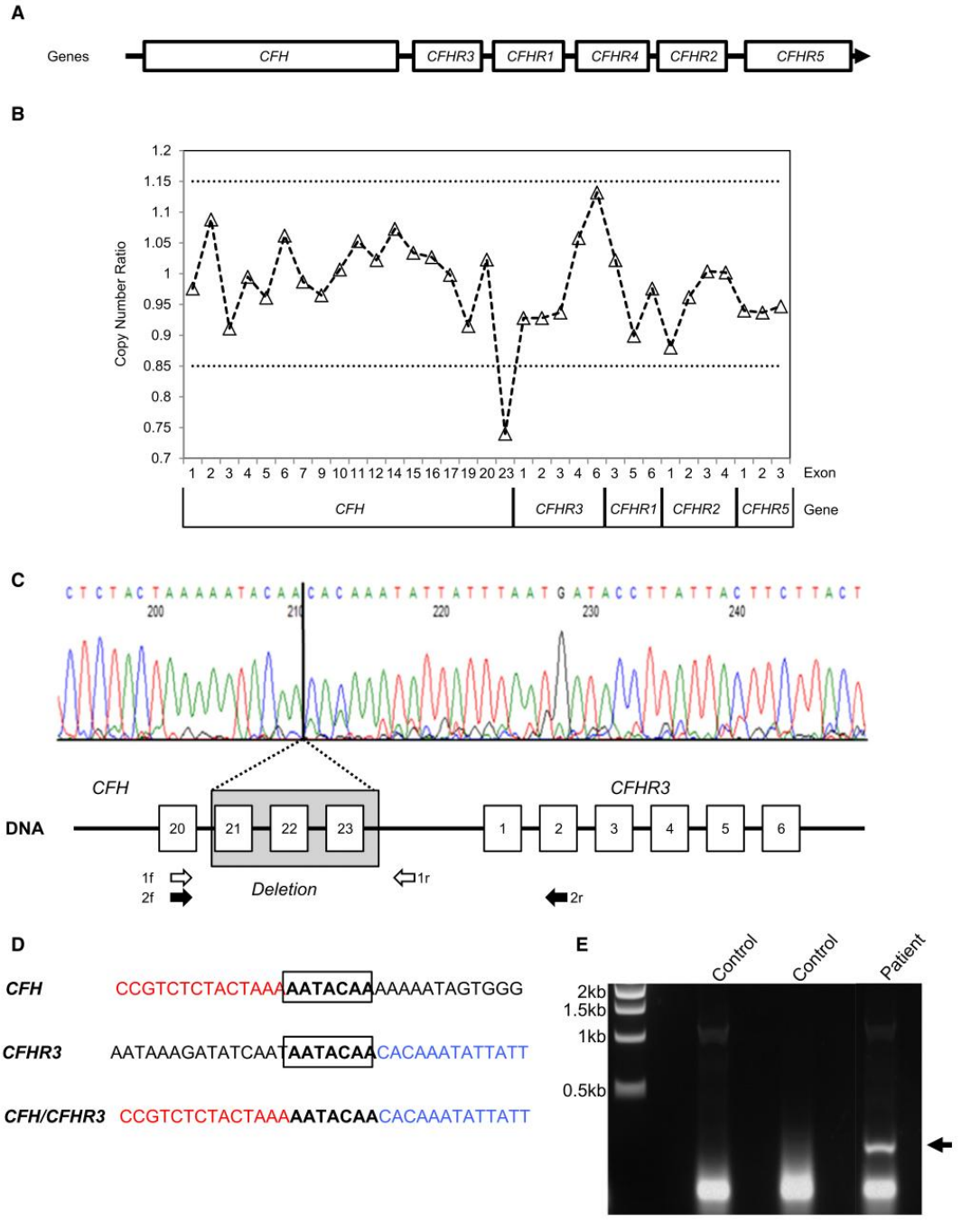

Figure 1. A 6.3-kb deletion in the RCA cluster results in a novel *CFH/CFHR3* hybrid gene transcript. (A) Genomic organization of the RCA cluster region containing the *CFH* and *CFHR* genes. (B) MLPA of *CFH*, *CFHR3*, *CFHR1*, *CFHR2*, and *CFHR5* showing the copy number ratio. The dotted lines represent ratios considered within the normal range. (C) Identification of breakpoint. Genomic DNA was amplified

In this study, we report a novel *CFH/CFHR3* hybrid gene arising through MMEJ that impairs cell surface complement regulation.

An 8-month-old boy presented with a diarrheal illness; on admission, creatinine was 52 μmol/L, and urea was 11.1 μmol/L. A blood film showed microangiopathic hemolytic anemia, and lactate dehydrogenase was 5747 units/L. Stool culture was positive for *Escherichia coli* O157:H7, and a diagnosis of Shiga toxin–producing *E. coli* (STEC) hemolytic uremic syndrome (HUS) was made. He did not require RRT and did not receive plasma exchange. He was discharged 2 weeks later with a creatinine of 107 μmol/L. He was readmitted 2 weeks later with an upper respiratory tract infection. Creatinine on readmission was 141 μmol/L, urea was 20.6 μmol/L, lactate dehydrogenase was 2398 units/L, and platelets were 128×10^9/L, with evidence of microangiopathic hemolytic anemia on blood film. This relapse suggested a diagnosis of aHUS rather than STEC HUS. Serum complement levels were within the normal range: C3, 1.05 g/L (0.68–1.80 g/L); C4, 0.30 g/L (0.18–0.60 g/L); and FH, 0.51 g/L (0.35–0.59 g/L). He commenced plasma exchange and required three sessions of hemofiltration. A renal biopsy confirmed a thrombotic microangiopathy (Figure 2). Creatinine returned to a baseline of approximately 100 μmol/L. He was treated with weekly plasma exchange for 3.5 years before starting Eculizumab, on which he has been maintained for 3 years. Creatinine is currently 200 μmol/L, and there have been no additional episodes of aHUS while on treatment with Eculizumab (Supplemental Figure 2).

Sanger sequencing of aHUS-associated genes (*CFH*, *CFI*, *CFB*, *CD46*, *C3*, *THBD*, and *DGKE*) did not reveal any mutations, although heterozygosity was found for the common Y402H polymorphism (rs1061170) in *CFH*. Multiplex ligation–dependent probe amplification (MLPA) of *CFH* and the *CFHR*s revealed a deletion in the *CFH* gene (Figure 1B). To define the extent of the deletion and the breakpoint, Sanger sequencing of genomic DNA was undertaken. This showed a 6.3-kb deletion extending from *CFH* intron 20 to the *CFH* 3′ intergenic region incorporating exons 21–23 of *CFH* (Figure 1C). Directly flanking the breakpoint was a 7-bp region of microhomology (Figure 1D). The deletion was not detected in either parent, suggesting that this was a *de novo* event.

We hypothesized that the loss of *CFH* exons 21–23 and the 3′ UTR of *CFH* would lead to aberrant splicing. *In silico* analysis showed that the next acceptor splice site following the *CFH* exon 20 donor site was 5′ of *CFHR3* exon 2, thus potentially producing a transcript for a hybrid *CFH/CFHR3* gene (Supplemental Figure 3). mRNA for this putative hybrid *CFH/CFHR3* gene was confirmed by amplifying patient and control cDNA with *CFH*– and *CFHR3*–specific primers. This showed a product in the patient and not in healthy controls (Figure 1E).

To confirm that this hybrid transcript led to the synthesis and secretion of a hybrid protein, Western blotting was performed using a series of epitope–defined anti–FH mAbs (Figure 3A). An initial blot probed with OX24, an FH CCP5–specific antibody, showed, in addition to FH, two additional bands only in the patient (Figure 3B). The species at approximately 160 kD was consistent with the predicted intact hybrid FH/FHR3. A species at approximately 120 kD was hypothesized to be a degradation product. No abnormal bands were detected in either parent (Supplemental Figure 4A). An FH C terminus–specific antibody (L20/3) did not reveal additional aberrant bands, indicating that these protein species lacked FH CCP20 (Figure 3C). Genotyping showed the patient to be heterozygous for the Y402H polymorphism in FH. This allowed the use of mAbs specific for histidine or tyrosine in CCP7 of FH.[14] The normal allele's product reacted to the histidine-specific mAb (MBI7) and gave a single band of approximately 150 kD, consistent with FH. The hybrid allele reacted to the tyrosine-specific mAb (MBI6), showing a doublet at approximately 160 kD (Figure 3D). FHR3 is known to be alternately glycosylated,[15–17] and we hypothesize that the doublet is caused by an alternately glycosylated hybrid protein. Additionally, it can be seen that the presumed degradation product at 120 kD is only seen using the tyrosine-specific mAb, consistent with this only arising from the hybrid protein.

To confirm these findings, FH species were purified from serum using affinity chromatography with an OX24 mAb. These were separated by a 6% SDS-PAGE before trypsin digestion. Mass spectrometry was then carried out on these three purified bands (Figure 4A). Peptide species identified from the approximately 160-kD band confirmed that this was a hybrid FH/FHR3 protein. Protein fragments from the 150-kD band were consistent with FH. Fragments from CCP5, CCP6, CCP8, CCP9, and CCP14 of FH were seen in the band at approximately 120 kD. Coverage at CCP7 was insufficient to determine whether there was a tyrosine or histidine at position 402. The Western blot analysis and mass spectrometry data together show that breakdown

using a forward primer specific for *CFH* exon 20 (1f) and a reverse primer in the *CFH* 3′ region intergenic region (1r) and sequenced. The deletion of exons 21–23 of *CFH* is shown (shaded box), and the breakpoint is highlighted. (D) The *CFH* and *CFHR3* sequence flanking the breakpoint. The 7-bp area of microhomology is shown (bold and boxed). (E) Confirmation of hybrid *CFH/CFHR3* mRNA. A message for the hybrid *CFH/CFHR3* gene was confirmed by amplifying patient and control cDNA with cross *CFH* and *CFHR3* gene primers (black arrows 2f and 2r in C). The agarose gel shows amplified cDNA. The patient lane shows an amplified product consistent with a hybrid *CFH/CFHR3* gene, which is not present in either control cDNA.

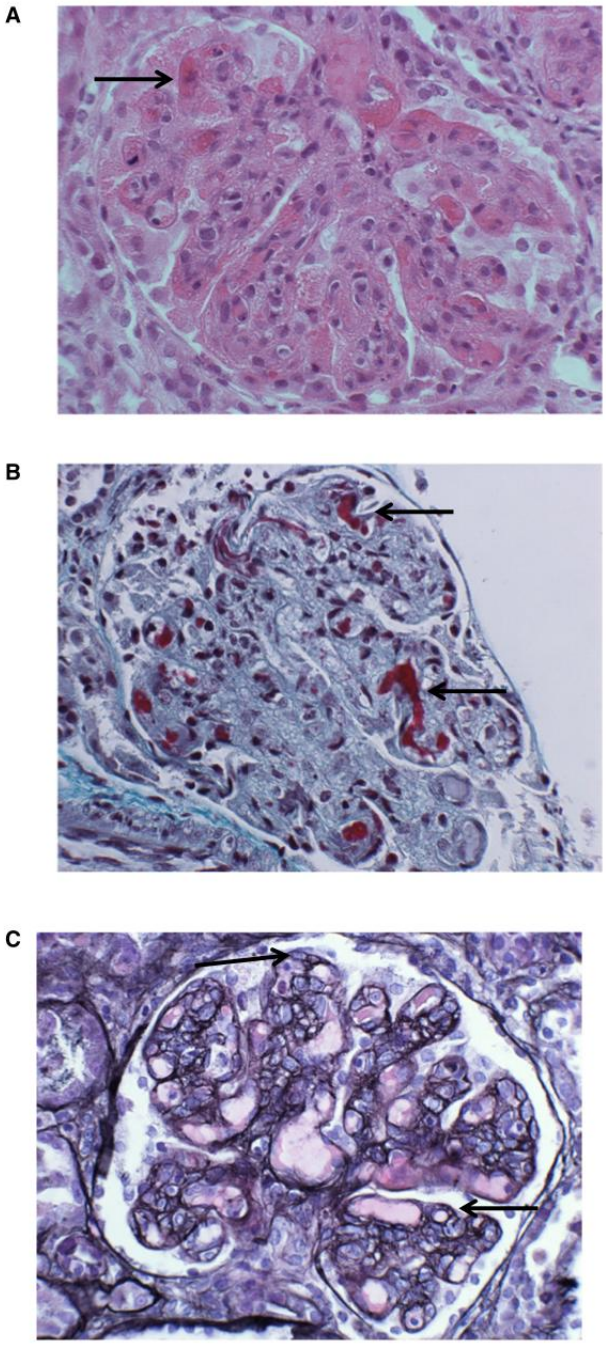

Figure 2. A renal biopsy from the patient demonstrating haemolytic uraemic syndrome. Renal biopsy of the patient showing thrombi (arrows) on capillary loops. (A) hematoxylin and eosin. (B) Masson trichrome. A developing membranoproliferative pattern of injury can be seen in (C) characterized by capillary loop double contours (arrows). Silver counterstained H&E, ×400.

products from only the hybrid protein are seen in serum.

To examine the functional significance, the FH/FHR3 hybrid protein was purified from serum using affinity chromatography with an MBI6 mAb followed by gel filtration (Supplemental Figure 4B). The FH/FHR3 hybrid protein displayed both impaired cell surface decay acceleration (Figure 4B) and cofactor activity (Figure 4C, Supplemental Figure 5).

Thus, in this study, we show a deletion in the RCA cluster resulting in a novel *de novo CFH/CFHR3* hybrid gene. Microhomology in the sequence flanking the breakpoint suggests that the deletion has occurred through MMEJ. Genomic disorders affecting *CFH* and the *CFHR*s as a result of nonallelic homologous recombination are found in approximately 4.5% of patients with aHUS. In contrast, only one patient with a genomic disorder secondary to MMEJ has been described.[11]

We have shown that the product of this gene, a 22 CCP domain protein, is secreted, albeit with degradation fragments present in the serum, suggesting impaired stability of the protein. Functional analysis of the purified FH/FHR3 protein showed impaired cell surface complement regulation.

This FH/FHR3 hybrid protein is similar to that described by Francis *et al.*,[11] in that in both, the C-terminal end of FH is replaced by all five CCPs of FHR3 (Supplemental Figure 1). Although both hybrids lack CCP20 of FH responsible for cell surface protection, the hybrid reported here also lacks CCP18 and CCP19.

The functional role of FHR3 is unclear, with various regulatory activities being suggested.[16,18] Unsurprisingly, given its high homology with CCP7 of FH, FHR3 binds to heparin.[16] FHR3 also binds to C3b and C3d through CCP4 and CCP5.[16,19] Competition between FHR3 and FH for surface-bound C3b has been described.[18] The hybrid described by Francis *et al.*[11] showed normal fluid–phase complement

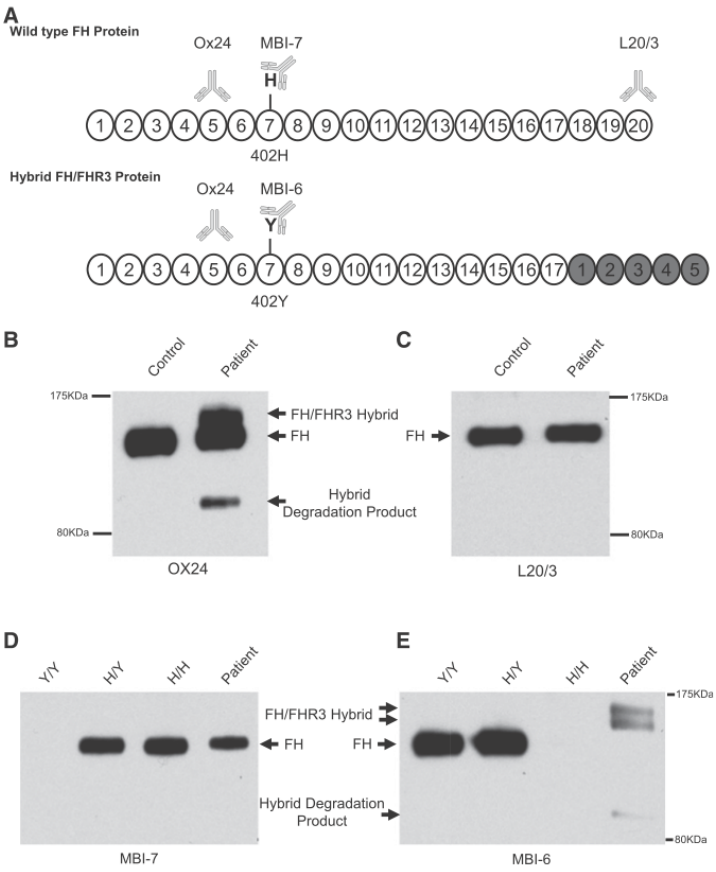

Figure 3. Identification of a novel FH/FHR3 hybrid protein in patient serum. (A) The protein product of wild-type FH and the predicted FH/FHR3 hybrid protein consisting of CCP1–17 of FH and CCP1–5 of FHR3. The CCPs originating from FHR3 are highlighted in gray. The patient is heterozygous for a common polymorphism in CCP7 of FH (Y402H). Binding epitopes for the mAbs are shown: Ox24-CCP5, MBI7-CCP7 amino acid 402 histidine variant, MBI6-CCP7 amino acid 402 tyrosine variant, and L20/3-CCP20. (B) Serum Western blot using OX24 showing two additional bands in the patient in addition to FH. The upper band is consistent with the predicted size of the FH/FHR3 hybrid. (C) Serum Western blot using L20/3. When using this FH C terminus–specific antibody, no additional bands are seen, consistent with the hybrid protein lacking CCP18–20. (D) Serum Western blot using MBI7. This shows that the patient has a normal allele producing FH with the histidine at amino acid 402. (E) Serum Western blot using MBI6. The patient has three additional bands and no wild–type FH band. This is consistent with the FH/FHR3 hybrid protein carrying the tyrosine amino acid at position 402. The two additional upper bands represent differentially glycosylated hybrid protein, consistent with that seen in the native FHR3. There is a faint degradation product only detected with the MBI6 antibody. This is consistent with the breakdown product being generated only from the hybrid protein and suggests a less stable protein product. The controls were unaffected, unrelated, genotyped samples. In B and C, a sample from an individual heterozygous at Y402H was used. In D and E, samples from individuals homozygous for Y402 (Y/Y), homozygous for H402 (H/H), or heterozygous (Y/H) were used.

regulatory activity but a profoundly reduced cell surface complement regulatory activity.

The fact that such an FH/FHR3 hybrid should lack cell surface regulatory activity is not surprising. Both of these hybrids lack the CCP20 domain of FH responsible for cell surface localization.[6] Additionally, structural analysis has revealed a specific hairpin structure suggesting a model whereby cell surface–bound C3b is engaged by both CCP1–4 and CCP19 and CCP20 of FH concurrently.[20] The elongation of FH by the addition of extra CCPs would prevent such an orientation.

Although the initial presentation of disease was seen in association with STEC, the rapid relapse suggested aHUS rather than STEC HUS, and this was confirmed by the finding of the *CFH/CFHR3* hybrid gene. In individuals carrying mutations in complement genes, additional triggers (*e.g.*, pregnancy[21] and infection[22]) are required for disease to manifest.[4,22] In this case, STEC served to unmask the genetic predisposition to disease, such as in two previously reported patients with STEC-triggered aHUS.[23]

In summary, we describe a deletion in the RCA cluster arising through MMEJ that results in a novel *CFH/CFHR3* hybrid gene. The fact that this arose as a *de novo* event suggests that this is a dynamic area of the genome where we should expect additional genomic disorders.

CONCISE METHODS

The study was approved by Newcastle and North Tyneside 1 Research Ethics Committee, and informed consent was obtained in accordance with the Declaration of Helsinki.

Complement Assays

C3 and C4 levels were measured by rate nephelometry (Beckman Coulter Array 360). FH levels were measured by radial immunodiffusion (Binding Site). Screening for FH autoantibodies was undertaken using ELISA as described previously.[24]

Genetic Analyses and MLPA

Mutation screening of *CFH*,[25] *CFI*,[26] *CFB*,[27] *MCP*,[28] *C3*,[29] and *DGKE*[30] was undertaken using Sanger sequencing as

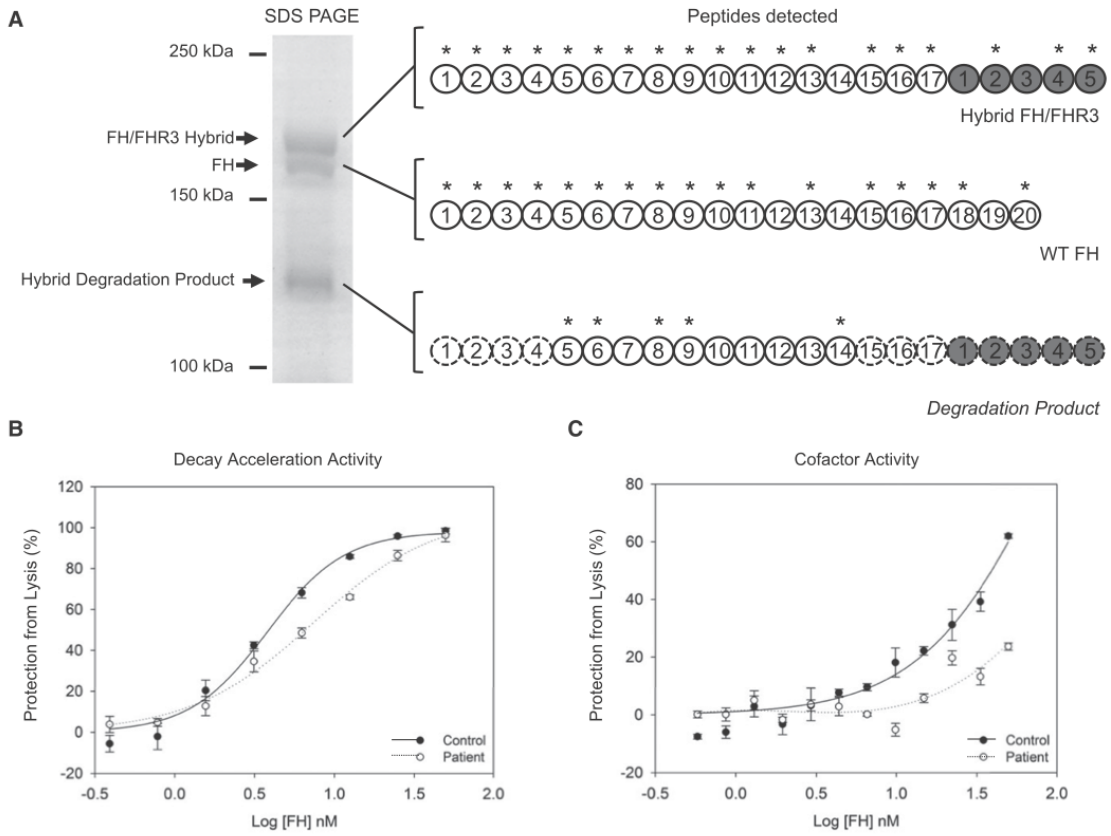

Figure 4. Impaired cell surface co-factor and decay acceleration activity of hybrid FH/FHR3 hybrid protein. (A) Mass spectrometry of purified proteins. The FH, FH/FHR3 hybrid, and degradation product were purified using affinity chromatography with an OX24 column. These FH species were separated by 6% SDS-PAGE, stained using Coomassie, cut from the gel (left panel), and submitted for trypsin digest and mass spectrometry. Peptides sequences identified using mass spectrometry are indicated with asterisks. The peptides detected in the hybrid degradation product by mass spectrometry are annotated on a full–length FH/FHR3 hybrid protein. CCPs that cannot be directly inferred from mass spectrometry data are outlined with a dashed line. A molecular mass of approximately 120 kD is consistent with an approximately 17 CCP protein. (B) Decay acceleration assays on sheep erythrocytes. The purified FH/FHR3 hybrid from the patient showed impaired cell surface complement regulation compared with wild type (WT) FH purified from control. Alternative pathway convertase (C3bBb) was formed on sheep erythrocytes. Cells were incubated for 15 minutes with dilutions of purified FH/FHR3 hybrid and WT FH before triggering lysis with NHSΔBΔH. Maximal lysis occurs in buffer-only (0 mM FH) conditions. Addition of WT FH caused decay of the C3 convertase and decay of convertase resulting in inhibition of lysis. The FH/FHR3 hybrid was up to 2-fold less efficient at inhibiting lysis. (C) C3b cofactor activity on sheep erythrocytes. WT FH and purified FH/FHR3 hybrid were tested for the ability to act as a cofactor for factor I–catalyzed inactivation of C3b deposited on the surfaces of sheep erythrocytes. The C3 convertase (C3bBb) was formed on residual C3b, and lysis was triggered by adding NHSΔBΔH. Maximal lysis occurs in the presence of buffer only (0 mM FH). The addition of WT FH and factor I produces iC3b, which decreases convertase formation and subsequent lysis, and this is shown as increasing amounts of inhibition of lysis (expressed as percentage of maximal lysis) after incubation with factor I and WT FH (black circles) or FH/FHR3 hybrid (white circles). The FH/FHR3 hybrid can be seen to be markedly less active than WT.

previously described. Screening for genomic disorders affecting *CFH*, *CFHR1*, *CFHR2*, *CFHR3*, and *CFHR5* was undertaken using MLPA in both the affected individual and 500 normal controls as previously described.[11] A proprietary kit from MRC Holland (SALSA MLPA Kit P236-A1 ARMD; www.mlpa.com) was supplemented by specific in–house *CFH* probes (Supplemental Material). GeneMarker software (Version 1.90) was used to calculate dosage quotients.

Genomic Breakpoint Analyses

To identify the breakpoint of the deletion that resulted in the *CFH/CFHR3* hybrid gene, genomic DNA was amplified using a forward primer specific for *CFH* in exon 20 and a reverse primer in the *CFH* 3′ region (Figure 1C). The sequence of the forward primer was GTAACTGTTATCAGTTGATTTGC, and the reverse was ACGGATTGCATGTATAAGTG. The product was then sequenced using direct fluorescent sequencing.

Confirmation of *CFH/CFHR3* RNA Product

mRNA was extracted from peripheral blood lymphocytes of the patient, and cDNA was prepared. This was amplified using a forward primer in *CFH* exon 20 (TGGATGGAGCCAGTAATGTAACATGCAT) and a reverse primer in *CFHR3* exon 2 (GAAATAGACCTCCATGTTTAATGTCTG) (Figure 1C). The PCR product was detected in an ethidium bromide–stained 1.5% agarose gel.

Western Blotting

Detection of abnormal protein products in serum arising as a consequence of the deletion was undertaken by Western blotting. Sera were diluted 1:100, and under nonreducing conditions, they were electrophoresed on 6% SDS-PAGE gel and transferred onto nitrocellulose. mAbs of known specificity to FH (OX24, CCP5, L20/3, CCP20, MBI6, CCP7 402Y, MBI7 CCP7, and 402H) were used with sheep anti–mouse Ig HRP (Jackson Immuno Research) at dilutions of 1:2000 (primary antibody) and 1:4000 (secondary antibody) (Supplemental Material). After washes in 1× TBST, blots were developed using Pierce ECL Western Blotting Substrate (Thermo Scientific).

Purification of FH Species

FH species (wild-type FH from normal allele, FH/FHR3 hybrid, and FH/FHR3 breakdown product all with CCP5 of FH) were purified from serum with affinity chromatography using immobilized mAb to FH (OX24) on a 1-ml HiTrap NHS HP Column (GE Healthcare). After washing with PBS, the bound proteins were then eluted using 0.1 M glycine (pH 2.7). The eluted material was pooled and concentrated for analysis by mass spectrometry.

The FH/FHR3 hybrid protein was purified from patient serum with affinity chromatography using immobilized mAb to FH CCP7 402Y (MBI6)[31] on a 1-ml HiTrap NHS HP Column (GE Healthcare). After washing with PBS, the bound proteins were eluted using PBS/diethylamine (50 mM). Subsequently, the FH/FHR3 protein was purified from its degradation product and FHL1 using a Superdex 200 Size Exclusion Column. We also purified wild–type FH protein from a healthy control. FH or FH/FHR3 hybrid protein purified using the MBI6 column was used for cofactor and decay acceleration assays.

Mass Spectrometry

A 6% SDS-PAGE was run, and the three bands identified by Coomassie staining were excised from the gel as indicated in Figure 4A. Trypsin digest and mass spectrometry were then undertaken (Supplemental Material).

Cell Surface Decay Acceleration Assays

Decay acceleration on sheep erythrocytes was performed as previously described[32] and in Supplemental Material. Briefly, FH/FHR3 hybrid from the patient and FH from controls were purified by immunoaffinity chromatography as above and gel filtered into PBS. Alternative pathway convertase (C3bBb) was formed on sheep erythrocytes. Cells were incubated for 15 minutes with 1.24–50 nM patient FH/FHR3 hybrid or control FH. The molar concentration of the purified patient FH/FHR3 was estimated using the extinction coefficient (272,170 M cm^{-1}), whereas the extinction coefficient of FH (246,800 M cm^{-1}) was used for the control. Lysis was initiated with 4% normal human serum depleted of factor B and FH (NHSΔBΔH). Maximal lysis was achieved by adding NHSΔBΔH to no FH wells (buffer only). To determine the amount of lysis, cells were pelleted by centrifugation, and hemoglobin release was measured at 420 nm (A_{420}). Percentage of inhibition of lysis in the presence of increasing concentrations was defined as (A_{420}[buffer only] − A_{420}[FH])/A_{420}[buffer only] ×100%.

Cell Surface Cofactor Assays

Cofactor activity on sheep erythrocytes was performed as previously described.[32] Briefly washed C3b–coated sheep erythrocyte cells were resuspended in AP buffer and incubated with an equal volume of a range of concentrations of FH/FHR3 hybrid and wild-type FH and 2.5 μg/ml factor I (CompTech) for 8 minutes at 25°C. After three washes in AP buffer, AP convertase was formed on the remaining C3b. Lysis was initiated with 4% NHSΔBΔH. Again, cells were pelleted, and hemoglobin release was measured at 420 nm. Percentage of inhibition from lysis was calculated by the formula (A_{420}[buffer only] − A_{420}[FH])/A_{420}[buffer only] ×100%.

ACKNOWLEDGMENTS

We thank Achim Treumann for technical help with mass spectrometry.

The research leading to these results has received funding from European Union's Seventh Framework Programme FP7/2007-2013 Grant 305608 (EURenOmics). Funding for this study was provided by United Kingdom Medical Research Council Grant G0701325. G.S.R.A. is funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico. E.K.S.W. is a Medical Research Council Clinical Training Fellow. D.K. is a Wellcome Trust Intermediate Clinical Fellow.

DISCLOSURES

A.A. has received fees from Alexion Pharmaceuticals for lectures. C.L.H. is also employed by GlaxoSmithKline and has shares in this company. Newcastle University has received fees from Alexion Pharmaceuticals for lectures and consultancy undertaken by T.H.J.G. and D.K.

REFERENCES

1. Díaz-Guillén MA, Rodríguez de Córdoba S, Heine-Suñer D: A radiation hybrid map of complement factor H and factor H-related genes. *Immunogenetics* 49: 549–552, 1999
2. Male DA, Ormsby RJ, Ranganathan S, Giannakis E, Gordon DL: Complement factor H: Sequence analysis of 221 kb of human genomic DNA containing the entire fH, fHR-1 and fHR-3 genes. *Mol Immunol* 37: 41–52, 2000
3. Zipfel PF, Jokiranta TS, Hellwage J, Koistinen V, Meri S: The factor H protein family. *Immunopharmacology* 42: 53–60, 1999
4. Kavanagh D, Goodship TH, Richards A: Atypical hemolytic uremic syndrome. *Semin Nephrol* 33: 508–530, 2013
5. Ferreira VP, Herbert AP, Cortés C, McKee KA, Blaum BS, Esswein ST, Uhrín D, Barlow PN, Pangburn MK, Kavanagh D: The binding of factor H to a complex of physiological polyanions and C3b on cells is impaired in atypical hemolytic uremic syndrome. *J Immunol* 182: 7009–7018, 2009

6. Schmidt CQ, Herbert AP, Kavanagh D, Gandy C, Fenton CJ, Blaum BS, Lyon M, Uhrín D, Barlow PN: A new map of glycosaminoglycan and C3b binding sites on factor H. *J Immunol* 181: 2610–2619, 2008
7. Blaum BS, Hannan JP, Herbert AP, Kavanagh D, Uhrín D, Stehle T: Structural basis for sialic acid-mediated self-recognition by complement factor H. *Nat Chem Biol* 11: 77–82, 2015
8. Heinen S, Sanchez-Corral P, Jackson MS, Strain L, Goodship JA, Kemp EJ, Skerka C, Jokiranta TS, Meyers K, Wagner E, Robitaille P, Esparza-Gordillo J, Rodriguez de Cordoba S, Zipfel PF, Goodship TH: De novo gene conversion in the RCA gene cluster (1q32) causes mutations in complement factor H associated with atypical hemolytic uremic syndrome. *Hum Mutat* 27: 292–293, 2006
9. Venables JP, Strain L, Routledge D, Bourn D, Powell HM, Warwicker P, Diaz-Torres ML, Sampson A, Mead P, Webb M, Pirson Y, Jackson MS, Hughes A, Wood KM, Goodship JA, Goodship TH: Atypical haemolytic uraemic syndrome associated with a hybrid complement gene. *PLoS Med* 3: e431, 2006
10. Maga TK, Meyer NC, Belsha C, Nishimura CJ, Zhang Y, Smith RJ: A novel deletion in the RCA gene cluster causes atypical hemolytic uremic syndrome. *Nephrol Dial Transplant* 26: 739–741, 2011
11. Francis NJ, McNicholas B, Awan A, Waldron M, Reddan D, Sadlier D, Kavanagh D, Strain L, Marchbank KJ, Harris CL, Goodship TH: A novel hybrid CFH/CFHR3 gene generated by a microhomology-mediated deletion in familial atypical hemolytic uremic syndrome. *Blood* 119: 591–601, 2012
12. Eyler SJ, Meyer NC, Zhang Y, Xiao X, Nester CM, Smith RJ: A novel hybrid CFHR1/CFH gene causes atypical hemolytic uremic syndrome. *Pediatr Nephrol* 28: 2221–2225, 2013
13. Valoti E, Alberti M, Tortajada A, Garcia-Fernandez J, Gastoldi S, Besso L, Bresin E, Remuzzi G, Rodriguez de Cordoba S, Noris M: A novel atypical hemolytic uremic syndrome-associated hybrid CFHR1/CFH gene encoding a fusion protein that antagonizes factor H-dependent complement regulation. *J Am Soc Nephrol* 26: 209–219, 2015
14. Hakobyan S, Tortajada A, Harris CL, de Córdoba SR, Morgan BP: Variant-specific quantification of factor H in plasma identifies null alleles associated with atypical hemolytic uremic syndrome. *Kidney Int* 78: 782–788, 2010
15. Skerka C, Kühn S, Günther K, Lingelbach K, Zipfel PF: A novel short consensus repeat-containing molecule is related to human complement factor H. *J Biol Chem* 268: 2904–2908, 1993
16. Hellwage J, Jokiranta TS, Koistinen V, Vaarala O, Meri S, Zipfel PF: Functional properties of complement factor H-related proteins FHR-3 and FHR-4: Binding to the C3d region of C3b and differential regulation by heparin. *FEBS Lett* 462: 345–352, 1999
17. Skerka C, Chen Q, Fremeaux-Bacchi V, Roumenina LT: Complement factor H related proteins (CFHRs). *Mol Immunol* 56: 170–180, 2013
18. Fritsche LG, Lauer N, Hartmann A, Stippa S, Keilhauer CN, Oppermann M, Pandey MK, Köhl J, Zipfel PF, Weber BH, Skerka C: An imbalance of human complement regulatory proteins CFHR1, CFHR3 and factor H influences risk for age-related macular degeneration (AMD). *Hum Mol Genet* 19: 4694–4704, 2010
19. Hellwage J, Jokiranta TS, Friese MA, Wolk TU, Kampen E, Zipfel PF, Meri S: Complement C3b/C3d and cell surface polyanions are recognized by overlapping binding sites on the most carboxyl-terminal domain of complement factor H. *J Immunol* 169: 6935–6944, 2002
20. Morgan HP, Schmidt CQ, Guariento M, Blaum BS, Gillespie D, Herbert AP, Kavanagh D, Mertens HD, Svergun DI, Johansson CM, Uhrín D, Barlow PN, Hannan JP: Structural basis for engagement by complement factor H of C3b on a self surface. *Nat Struct Mol Biol* 18: 463–470, 2011
21. Fakhouri F, Roumenina L, Provot F, Sallée M, Caillard S, Couzi L, Essig M, Ribes D, Dragon-Durey MA, Bridoux F, Rondeau E, Frémeaux-Bacchi V: Pregnancy-associated hemolytic uremic syndrome revisited in the era of complement gene mutations. *J Am Soc Nephrol* 21: 859–867, 2010
22. Caprioli J, Noris M, Brioschi S, Pianetti G, Castelletti F, Bettinaglio P, Mele C, Bresin E, Cassis L, Gamba S, Porrati F, Bucchioni S, Monteferrante G, Fang CJ, Liszewski MK, Kavanagh D, Atkinson JP, Remuzzi G; International Registry of Recurrent and Familial HUS/TTP: Genetics of HUS: The impact of MCP, CFH, and IF mutations on clinical presentation, response to treatment, and outcome. *Blood* 108: 1267–1279, 2006
23. Alberti M, Valoti E, Piras R, Bresin E, Galbusera M, Tripodo C, Thaiss F, Remuzzi G, Noris M: Two patients with history of STEC-HUS, posttransplant recurrence and complement gene mutations. *Am J Transplant* 13: 2201–2206, 2013
24. Moore I, Strain L, Pappworth I, Kavanagh D, Barlow PN, Herbert AP, Schmidt CQ, Staniforth SJ, Holmes LV, Ward R, Morgan L, Goodship TH, Marchbank KJ: Association of factor H autoantibodies with deletions of CFHR1, CFHR3, CFHR4, and with mutations in CFH, CFI, CD46, and C3 in patients with atypical hemolytic uremic syndrome. *Blood* 115: 379–387, 2010
25. Richards A, Buddles MR, Donne RL, Kaplan BS, Kirk E, Venning MC, Tielemans CL, Goodship JA, Goodship TH: Factor H mutations in hemolytic uremic syndrome cluster in exons 18-20, a domain important for host cell recognition. *Am J Hum Genet* 68: 485–490, 2001
26. Kavanagh D, Kemp EJ, Mayland E, Winney RJ, Duffield JS, Warwick G, Richards A, Ward R, Goodship JA, Goodship TH: Mutations in complement factor I predispose to development of atypical hemolytic uremic syndrome. *J Am Soc Nephrol* 16: 2150–2155, 2005
27. Kavanagh D, Kemp EJ, Richards A, Burgess RM, Mayland E, Goodship JA, Goodship TH: Does complement factor B have a role in the pathogenesis of atypical HUS? *Mol Immunol* 43: 856–859, 2006
28. Richards A, Kemp EJ, Liszewski MK, Goodship JA, Lampe AK, Decorte R, Müslümanoğlu MH, Kavukcu S, Filler G, Pirson Y, Wen LS, Atkinson JP, Goodship TH: Mutations in human complement regulator, membrane cofactor protein (CD46), predispose to development of familial hemolytic uremic syndrome. *Proc Natl Acad Sci U S A* 100: 12966–12971, 2003
29. Frémeaux-Bacchi V, Miller EC, Liszewski MK, Strain L, Blouin J, Brown AL, Moghal N, Kaplan BS, Weiss RA, Lhotta K, Kapur G, Mattoo T, Nivet H, Wong W, Gie S, Hurault de Ligny B, Fischbach M, Gupta R, Hauhart R, Meunier V, Loirat C, Dragon-Durey MA, Fridman WH, Janssen BJ, Goodship TH, Atkinson JP: Mutations in complement C3 predispose to development of atypical hemolytic uremic syndrome. *Blood* 112: 4948–4952, 2008
30. Lemaire M, Frémeaux-Bacchi V, Schaefer F, Choi M, Tang WH, Le Quintrec M, Fakhouri F, Taque S, Nobili F, Martinez F, Ji W, Overton JD, Mane SM, Nürnberg G, Altmüller J, Thiele H, Morin D, Deschenes G, Baudouin V, Llanas B, Collard L, Majid MA, Simkova E, Nürnberg P, Rioux-Leclerc N, Moeckel GW, Gubler MC, Hwa J, Loirat C, Lifton RP: Recessive mutations in DGKE cause atypical hemolytic-uremic syndrome. *Nat Genet* 45: 531–536, 2013
31. Hakobyan S, Harris CL, Tortajada A, Goicochea de Jorge E, García-Layana A, Fernández-Robredo P, Rodríguez de Córdoba S, Morgan BP: Measurement of factor H variants in plasma using variant-specific monoclonal antibodies: Application to assessing risk of age-related macular degeneration. *Invest Ophthalmol Vis Sci* 49: 1983–1990, 2008
32. Wong EK, Anderson HE, Herbert AP, Challis RC, Brown P, Reis GS, Tellez JO, Strain L, Fluck N, Humphrey A, Macleod A, Richards A, Ahlert D, Santibanez-Koref M, Barlow PN, Marchbank KJ, Harris CL, Goodship TH, Kavanagh D: Characterization of a factor H mutation that perturbs the alternative pathway of complement in a family with membranoproliferative GN. *J Am Soc Nephrol* 25: 2425–2433, 2014

This article contains supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2015010100/-/DCSupplemental.

Chapter 8: References

1000 Genomes website [Online]. Available: http://www.1000genomes.org/about [Accessed 05.09.12.

Analyzer Splice Tool website [Online]. Available: http://ibis.tau.ac.il/ssat/SpliceSiteFrame.htm [Accessed 01.10.14.

dbSNP website [Online]. Available: http://www.ncbi.nlm.nih.gov/SNP/ [Accessed 29.01.15.

Exome Variant Server website [Online]. Seattle, WA: NHLBI GO Exome Sequencing Project (ESP). Available: http://evs.gs.washington.edu/EVS/ [Accessed 29.01.15.

FATHMM website [Online]. Available: http://fathmm.biocompute.org.uk/index.html [Accessed 23.03.15.

Genome Browser website [Online]. Available: http://genome-euro.ucsc.edu/cgi-bin/hgGateway [Accessed 23.05.15.

GERP++ website [Online]. Available: http://mendel.stanford.edu/SidowLab/downloads/gerp/ [Accessed 29.01.15.

Ingenuity® Variant Analysis™ website [Online]. Available: http://www.qiagen.com/ingenuity [Accessed 30.09.14.

Integrative Genomics Viewer website [Online]. Available: http://www.broadinstitute.org/igv/ [Accessed 29.01.15.

Mutation Assessor website [Online]. Available: http://mutationassessor.org/ [Accessed 23.03.15.

Mutation Taster website [Online]. Available: http://www.mutationtaster.org/ [Accessed 29.01.15.

Pfam website [Online]. Available: http://pfam.xfam.org/ [Accessed 02.05.15.

PhyloP website [Online]. Available: http://compgen.bscb.cornell.edu/phast/help-pages/phyloP.txt [Accessed 05.09.12.

Phyre2 website [Online]. Available: http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index [Accessed 29.01.15.

PolyPhen-2 website [Online]. Available: http://genetics.bwh.harvard.edu/pph2/ [Accessed 29.01.15.

Protein Data Bank website [Online]. Available: http://www.rcsb.org/pdb/home/home.do [Accessed 29.01.15.

ProteoWizard website [Online]. Available: proteowizard.sourceforge.net [Accessed 24.03.15.

UCSC In-Silico PCR website [Online]. UCSC. Available: https://genome.ucsc.edu/cgi-bin/hgPcr?hgsid=392883337_XIXlcRLMW6kC6Ak0eQDs9N3Xuy3z.

Vertebrate Multiz Alignment & Conservation track website [Online]. UCSC. Available: http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=309786867&c=chr21&g=cons46way - a_cfg_phyloP [Accessed 10.03.15.

ABARRATEGUI-GARRIDO, C., MARTÍNEZ-BARRICARTE, R., LÓPEZ-TRASCASA, M., RODRÍGUEZ DE CÓRDOBA, S. & SÁNCHEZ-CORRAL, P. 2009. Characterization of complement factor H–related (CFHR) proteins in plasma reveals novel genetic variations of CFHR1 associated with atypical hemolytic uremic syndrome. *Blood,* 114**,** 4261-4271.

ABARRATEGUI-GARRIDO, C., MELGOSA, M., PENA-CARRION, A., DE JORGE, E. G., DE CORDOBA, S. R., LOPEZ-TRASCASA, M. & SANCHEZ-CORRAL, P. 2008. Mutations in proteins of the alternative pathway of complement and the pathogenesis of atypical hemolytic uremic syndrome. *American Journal of Kidney Diseases,* 52**,** 171-80.

ABBOTT, C., WEST, L., POVEY, S., JEREMIAH, S., MURAD, Z., DISCIPIO, R. & FEY, G. 1989. The gene for human complement component C9 mapped to chromosome 5 by polymerase chain reaction. *Genomics,* 4**,** 606-9.

ABECASIS, G. R., AUTON, A., BROOKS, L. D., DEPRISTO, M. A., DURBIN, R. M., HANDSAKER, R. E., KANG, H. M., MARTH, G. T. & MCVEAN, G. A. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491**,** 56-65.

ADAMS, E. M., BROWN, M. C., NUNGE, M., KRYCH, M. & ATKINSON, J. P. 1991. Contribution of the repeating domains of membrane cofactor protein (CD46) of the complement system to ligand binding and cofactor activity. *J Immunol,* 147**,** 3005-11.

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nat Methods,* 7**,** 248-9.

AGUIAR, C. L. & ERKAN, D. 2013. Catastrophic antiphospholipid syndrome: how to diagnose a rare but highly fatal disease. *Ther Adv Musculoskelet Dis,* 5**,** 305-14.

AIRD, D., ROSS, M. G., CHEN, W. S., DANIELSSON, M., FENNELL, T., RUSS, C., JAFFE, D. B., NUSBAUM, C. & GNIRKE, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol,* 12**,** R18.

AL-NOURI, Z. L., REESE, J. A., TERRELL, D. R., VESELY, S. K. & GEORGE, J. N. 2015. Drug-induced thrombotic microangiopathy: a systematic review of published reports. *Blood,* 125**,** 616-8.

ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K. & WALTER, P. 2002. The Self-Assembly and Dynamic Structure of Cytoskeletal Filaments. *Molecular Biology of the Cell.* 4th ed. NY, USA: Garland Science.

ANASTASIOU, G., GIALERAKI, A., MERKOURI, E., POLITOU, M. & TRAVLOU, A. 2012. Thrombomodulin as a regulator of the anticoagulant pathway: implication in the development of thrombosis. *Blood Coagul Fibrinolysis,* 23**,** 1-10.

ANDREWS, S. *FastQC A Quality Control tool for High Throughput Sequence Data* [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed 18.02.15.

BAILEY, J. A., GU, Z., CLARK, R. A., REINERT, K., SAMONTE, R. V., SCHWARTZ, S., ADAMS, M. D., MYERS, E. W., LI, P. W. & EICHLER, E. E. 2002. Recent Segmental Duplications in the Human Genome. *Science,* 297**,** 1003-1007.

BAMSHAD, M. J., NG, S. B., BIGHAM, A. W., TABOR, H. K., EMOND, M. J., NICKERSON, D. A. & SHENDURE, J. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics,* 12**,** 745-755.

BANDYOPADHYAY, M. & ROHRER, B. 2012. Matrix Metalloproteinase Activity Creates Pro-Angiogenic Environment in Primary Human Retinal Pigment Epithelial Cells Exposed to Complement. *Investigative Ophthalmology & Visual Science,* 53**,** 1953-1961.

BARBOUR, T., JOHNSON, S., COHNEY, S. & HUGHES, P. 2012. Thrombotic microangiopathy and associated renal disorders. *Nephrol Dial Transplant,* 27**,** 2673-85.

BARILLA-LABARCA, M. L., LISZEWSKI, M. K., LAMBRIS, J. D., HOURCADE, D. & ATKINSON, J. P. 2002. Role of membrane cofactor protein (CD46) in regulation of C4b and C3b deposited on cells. *J Immunol,* 168**,** 6298-304.

BARUA, M., BROWN, E. J., CHAROONRATANA, V. T., GENOVESE, G., SUN, H. & POLLAK, M. R. 2013. Mutations in the INF2 gene account for a significant proportion of familial but not sporadic focal and segmental glomerulosclerosis. *Kidney International,* 83**,** 316-22.

BENZAQUEN, L. R., NICHOLSON-WELLER, A. & HALPERIN, J. A. 1994. Terminal complement proteins C5b-9 release basic fibroblast growth factor and platelet-derived growth factor from endothelial cells. *J Exp Med,* 179**,** 985-92.

BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N. & BOURNE, P. E. 2000. The Protein Data Bank. *Nucleic Acids Res,* 28**,** 235-42.

BERTRAM, J. F., DOUGLAS-DENTON, R. N., DIOUF, B., HUGHSON, M. D. & HOY, W. E. 2011. Human nephron number: implications for health and disease. *Pediatr Nephrol,* 26**,** 1529-33.

BESBAS, N., KARPMAN, D., LANDAU, D., LOIRAT, C., PROESMANS, W., REMUZZI, G., RIZZONI, G., TAYLOR, C. M., VAN DE KAR, N. & ZIMMERHACKL, L. B. 2006. A classification of hemolytic uremic syndrome and thrombotic thrombocytopenic purpura and related disorders. *Kidney Int,* 70**,** 423-431.

BIENAIME, F., DRAGON-DUREY, M.-A., REGNIER, C. H., NILSSON, S. C., KWAN, W. H., BLOUIN, J., JABLONSKI, M., RENAULT, N., RAMEIX-WELTI, M.-A., LOIRAT, C., SAUTES-FRIDMAN, C., VILLOUTREIX, B. O., BLOM, A. M. & FREMEAUX-BACCHI, V. 2009. Mutations in components of complement influence the outcome of Factor I-associated atypical hemolytic uremic syndrome. *Kidney Int,* 77**,** 339-349.

BISHOP, D. T. 1994. Linkage Analysis: Progress and Problems. *Philosophical Transactions: Biological Sciences,* 344**,** 337-343.

BLACKMORE, T. K., SADLON, T. A., WARD, H. M., LUBLIN, D. M. & GORDON, D. L. 1996. Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H. *J Immunol,* 157**,** 5422-7.

BLANCHETTE, M., KENT, W. J., RIEMER, C., ELNITSKI, L., SMIT, A. F., ROSKIN, K. M., BAERTSCH, R., ROSENBLOOM, K., CLAWSON, H., GREEN, E. D., HAUSSLER, D. & MILLER, W. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res,* 14**,** 708-15.

BLAUM, B. S., HANNAN, J. P., HERBERT, A. P., KAVANAGH, D., UHRIN, D. & STEHLE, T. 2015. Structural basis for sialic acid-mediated self-recognition by complement factor H. *Nat Chem Biol,* 11**,** 77-82.

BLOM, A. M., KASK, L. & DAHLBÄCK, B. 2003. CCP1–4 of the C4b-binding protein α-chain are required for factor I mediated cleavage of complement factor C3b. *Molecular Immunology,* 39**,** 547-556.

BOHANA-KASHTAN, O., ZIPOREN, L., DONIN, N., KRAUS, S. & FISHELSON, Z. 2004. Cell signals transduced by complement. *Molecular Immunology,* 41**,** 583-597.

BOSE, B. & CATTRAN, D. 2014. Glomerular diseases: FSGS. *Clin J Am Soc Nephrol,* 9**,** 626-32.

BOTSTEIN, D. & RISCH, N. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet,* 33 Suppl**,** 228-37.

BOYCOTT, K. M., VANSTONE, M. R., BULMAN, D. E. & MACKENZIE, A. E. 2013. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet,* 14**,** 681-91.

BOYER, O., BENOIT, G., GRIBOUVAL, O., NEVO, F., TETE, M. J., DANTAL, J., GILBERT-DUSSARDIER, B., TOUCHARD, G., KARRAS, A., PRESNE, C., GRUNFELD, J. P., LEGENDRE, C., JOLY, D., RIEU, P., MOHSIN, N., HANNEDOUCHE, T., MOAL, V., GUBLER, M. C., BROUTIN, I., MOLLET, G. & ANTIGNAC, C. 2011a. Mutations in INF2 are a major cause of autosomal dominant focal segmental glomerulosclerosis. *Journal of the American Society of Nephrology,* 22**,** 239-45.

BOYER, O., NEVO, F., PLAISIER, E., FUNALOT, B., GRIBOUVAL, O., BENOIT, G., CONG, E. H., ARRONDEL, C., TETE, M. J., MONTJEAN, R., RICHARD, L., KARRAS, A., POUTEIL-NOBLE, C., BALAFREJ, L., BONNARDEAUX, A., CANAUD, G., CHARASSE, C., DANTAL, J., DESCHENES, G., DETEIX, P., DUBOURG, O., PETIOT, P., POUTHIER, D., LEGUERN, E., GUIOCHON-MANTEL, A., BROUTIN, I., GUBLER, M. C., SAUNIER, S., RONCO, P., VALLAT, J. M., ALONSO, M. A., ANTIGNAC, C. & MOLLET, G. 2011b. INF2 mutations in Charcot-Marie-Tooth disease with glomerulopathy. *N Engl J Med,* 365**,** 2377-88.

BOYER, O. & NIAUDET, P. 2011. Hemolytic uremic syndrome: new developments in pathogenesis and treatment. *Int J Nephrol,* 2011**,** 908407.

BRANDT, D., GIMONA, M., HILLMANN, M., HALLER, H. & MISCHAK, H. 2002. Protein kinase C induces actin reorganization via a Src- and Rho-dependent pathway. *J Biol Chem,* 277**,** 20903-10.

BREITSPRECHER, D. & FAIX, J. 2010. The inverted Formin INF2 sorts it out. *Dev Cell,* 18**,** 689-90.

BREITSPRECHER, D. & GOODE, B. L. 2013. Formins at a glance. *J Cell Sci,* 126**,** 1-7.

BRESIN, E., DAINA, E., NORIS, M., CASTELLETTI, F., STEFANOV, R., HILL, P., GOODSHIP, T. H. & REMUZZI, G. 2006. Outcome of renal transplantation in patients with non-Shiga toxin-associated hemolytic uremic syndrome: prognostic significance of genetic background. *Clin J Am Soc Nephrol,* 1**,** 88-99.

BROWN, E. J., SCHLONDORFF, J. S., BECKER, D. J., TSUKAGUCHI, H., TONNA, S. J., USCINSKI, A. L., HIGGS, H. N., HENDERSON, J. M. & POLLAK, M. R. 2010. Mutations in the formin gene INF2 cause focal segmental glomerulosclerosis. *Nature Genetics,* 42**,** 72-6.

BRUNEAU, S., NEEL, M., ROUMENINA, L. T., FRIMAT, M., LAURENT, L., FREMEAUX-BACCHI, V. & FAKHOURI, F. 2014. Loss of DGK induces endothelial cell activation and death independently of complement activation. *Blood*.

BU, F., MAGA, T., MEYER, N. C., WANG, K., THOMAS, C. P., NESTER, C. M. & SMITH, R. J. 2014. Comprehensive genetic analysis of complement and coagulation genes in atypical hemolytic uremic syndrome. *J Am Soc Nephrol,* 25**,** 55-64.

BUBECK, D. 2014. The making of a macromolecular machine: assembly of the membrane attack complex. *Biochemistry,* 53**,** 1908-15.

CAMILLERI, R. S., COHEN, H., MACKIE, I. J., SCULLY, M., STARKE, R. D., CRAWLEY, J. T., LANE, D. A. & MACHIN, S. J. 2008. Prevalence of the ADAMTS-13 missense mutation R1060W in late onset adult thrombotic thrombocytopenic purpura. *J Thromb Haemost,* 6**,** 331-8.

CAMILLERI, R. S., SCULLY, M., THOMAS, M., MACKIE, I. J., LIESNER, R., CHEN, W. J., MANNS, K. & MACHIN, S. J. 2012. A phenotype–genotype correlation of ADAMTS13 mutations in congenital thrombotic thrombocytopenic purpura patients treated in the United Kingdom. *Journal of Thrombosis and Haemostasis,* 10**,** 1792-1801.

CAMPBELL, R. D. 1987. The molecular genetics and polymorphism of C2 and factor B. *Br Med Bull,* 43**,** 37-49.

CAMPBELL, R. D., LAW, S. K., REID, K. B. & SIM, R. B. 1988. Structure, organization, and regulation of the complement genes. *Annu Rev Immunol,* 6**,** 161-95.

CAMPBELL, R. D. & PORTER, R. R. 1983. Molecular cloning and characterization of the gene coding for human complement protein factor B. *Proc Natl Acad Sci U S A,* 80**,** 4464-8.

CAMPELLONE, K. G. & WELCH, M. D. 2010. A nucleator arms race: cellular control of actin assembly. *Nat Rev Mol Cell Biol,* 11**,** 237-51.

CAPRIOLI, J., BETTINAGLIO, P., ZIPFEL, P. F., AMADEI, B., DAINA, E., GAMBA, S., SKERKA, C., MARZILIANO, N., REMUZZI, G. & NORIS, M. 2001. The molecular basis of familial hemolytic uremic syndrome: mutation analysis of factor H gene reveals a hot spot in short consensus repeat 20. *Journal of the American Society of Nephrology,* 12**,** 297-307.

CAPRIOLI, J., CASTELLETTI, F., BUCCHIONI, S., BETTINAGLIO, P., BRESIN, E., PIANETTI, G., GAMBA, S., BRIOSCHI, S., DAINA, E., REMUZZI, G. & NORIS, M. 2003. Complement factor H mutations and gene polymorphisms in haemolytic uraemic syndrome: the C-257T, the A2089G and the G2881T polymorphisms are strongly associated with the disease. *Human Molecular Genetics,* 12**,** 3385-95.

CAPRIOLI, J., NORIS, M., BRIOSCHI, S., PIANETTI, G., CASTELLETTI, F., BETTINAGLIO, P., MELE, C., BRESIN, E., CASSIS, L., GAMBA, S., PORRATI, F., BUCCHIONI, S., MONTEFERRANTE, G., FANG, C. J., LISZEWSKI, M. K., KAVANAGH, D., ATKINSON, J. P. & REMUZZI, G. 2006. Genetics of HUS: the impact of MCP, CFH, and IF mutations on clinical presentation, response to treatment, and outcome. *Blood,* 108**,** 1267-79.

CAREW, M. A., PALEOLOG, E. M. & PEARSON, J. D. 1992. The roles of protein kinase C and intracellular Ca2+ in the secretion of von Willebrand factor from human vascular endothelial cells. *Biochem J,* 286 (Pt 2)**,** 631-5.

CARMEL, I., TAL, S., VIG, I. & AST, G. 2004. Comparative analysis detects dependencies among the 5' splice-site positions. *RNA,* 10**,** 828-40.

CARNEY, D. F., LANG, T. J. & SHIN, M. L. 1990. Multiple signal messengers generated by terminal complement complexes and their role in terminal complement complex elimination. *J Immunol,* 145**,** 623-9.

CATTANEO, M. 1999. Hyperhomocysteinemia, atherosclerosis and thrombosis. *Thromb Haemost,* 81**,** 165-76.

CHALLIS, R. C., ARAUJO, G. S. R., WONG, E. K. S., ANDERSON, H. E., AWAN, A., DORMAN, A. M., WALDRON, M., WILSON, V., BROCKLEBANK, V., STRAIN, L., MORGAN, B. P., HARRIS, C. L., MARCHBANK, K. J., GOODSHIP, T. H. J. & KAVANAGH, D. 2015. A De Novo Deletion in the Regulators of Complement Activation Cluster Producing a Hybrid Complement Factor H/Complement Factor H–Related 3 Gene in Atypical Hemolytic Uremic Syndrome. *Journal of the American Society of Nephrology*.

CHAMBERS, M. C., MACLEAN, B., BURKE, R., AMODEI, D., RUDERMAN, D. L., NEUMANN, S., GATTO, L., FISCHER, B., PRATT, B., EGERTSON, J., HOFF, K., KESSNER, D., TASMAN, N., SHULMAN, N., FREWEN, B., BAKER, T. A., BRUSNIAK, M.-Y., PAULSE, C., CREASY, D., FLASHNER, L., KANI, K., MOULDING, C., SEYMOUR, S. L., NUWAYSIR, L. M., LEFEBVRE, B., KUHLMANN, F., ROARK, J., RAINER, P., DETLEV, S., HEMENWAY, T., HUHMER, A., LANGRIDGE, J., CONNOLLY, B., CHADICK, T., HOLLY, K., ECKELS, J., DEUTSCH, E. W., MORITZ, R. L., KATZ, J. E., AGUS, D. B., MACCOSS, M., TABB, D. L. & MALLICK, P. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotech,* 30**,** 918-920.

CHEN, J. M., COOPER, D. N., CHUZHANOVA, N., FEREC, C. & PATRINOS, G. P. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet,* 8**,** 762-75.

CHEN, J. M., COOPER, D. N., FEREC, C., KEHRER-SAWATZKI, H. & PATRINOS, G. P. 2010. Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol,* 20**,** 222-33.

CHESARONE, M. A., DUPAGE, A. G. & GOODE, B. L. 2010. Unleashing formins to remodel the actin and microtubule cytoskeletons. *Nat Rev Mol Cell Biol,* 11**,** 62-74.

CHHABRA, E. S. & HIGGS, H. N. 2006. INF2 Is a WASP homology 2 motif-containing formin that severs actin filaments and accelerates both polymerization and depolymerization. *J Biol Chem,* 281**,** 26754-67.

CHHABRA, E. S. & HIGGS, H. N. 2007. The many faces of actin: matching assembly factors with cellular structures. *Nat Cell Biol,* 9**,** 1110-1121.

CHHABRA, E. S., RAMABHADRAN, V., GERBER, S. A. & HIGGS, H. N. 2009. INF2 is an endoplasmic reticulum-associated formin protein. *J Cell Sci,* 122**,** 1430-40.

CHOI, M., SCHOLL, U. I., JI, W., LIU, T., TIKHONOVA, I. R., ZUMBO, P., NAYIR, A., BAKKALOGLU, A., OZEN, S., SANJAD, S., NELSON-WILLIAMS, C., FARHI, A., MANE, S. & LIFTON, R. P. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A,* 106**,** 19096-101.

CLARK, M. J., CHEN, R., LAM, H. Y. K., KARCZEWSKI, K. J., CHEN, R., EUSKIRCHEN, G., BUTTE, A. J. & SNYDER, M. 2011. Performance comparison of exome DNA sequencing technologies. *Nat Biotech,* 29**,** 908-914.

COCHRAN, J. B., PANZARINO, V. M., MAES, L. Y. & TECKLENBURG, F. W. 2004. Pneumococcus-induced T-antigen activation in hemolytic uremic syndrome and anemia. *Pediatric Nephrology,* 19**,** 317-21.

COLE, D. S. & MORGAN, B. P. 2003. Beyond lysis: how complement influences cell fate. *Clin Sci (Lond),* 104**,** 455-66.

CONSTANTINESCU, A. R., BITZAN, M., WEISS, L. S., CHRISTEN, E., KAPLAN, B. S., CNAAN, A. & TRACHTMAN, H. 2004. Non-enteropathic hemolytic uremic syndrome: causes and short-term course. *American Journal of Kidney Diseases,* 43**,** 976-82.

COPPO, P. & VEYRADIER, A. 2009. Thrombotic microangiopathies: towards a pathophysiology-based classification. *Cardiovasc Hematol Disord Drug Targets,* 9**,** 36-50.

CORNEC-LE GALL, E., DELMAS, Y., DE PARSCAU, L., DOUCET, L., OGIER, H., BENOIST, J.-F., FREMEAUX-BACCHI, V. & LE MEUR, Y. 2014. Adult-Onset Eculizumab-Resistant Hemolytic Uremic Syndrome Associated With Cobalamin C Deficiency. *American Journal of Kidney Diseases,* 63**,** 119-123.

COUZI, L., CONTIN-BORDES, C., MARLIOT, F., SARRAT, A., GRIMAL, P., MOREAU, J. F., MERVILLE, P. & FREMEAUX-BACCHI, V. 2008. Inherited deficiency of membrane cofactor protein expression and varying manifestations of

recurrent atypical hemolytic uremic syndrome in a sibling pair. *Am J Kidney Dis,* 52**,** e5-9.

CRAIG, R. & BEAVIS, R. C. 2004. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics,* 20**,** 1466-7.

CRAWLEY, J. T. & SCULLY, M. A. 2013. Thrombotic thrombocytopenic purpura: basic pathophysiology and therapeutic strategies. *Hematology Am Soc Hematol Educ Program,* 2013**,** 292-9.

CUNNINGHAM, F., AMODE, M. R., BARRELL, D., BEAL, K., BILLIS, K., BRENT, S., CARVALHO-SILVA, D., CLAPHAM, P., COATES, G., FITZGERALD, S., GIL, L., GIRÓN, C. G., GORDON, L., HOURLIER, T., HUNT, S. E., JANACEK, S. H., JOHNSON, N., JUETTEMANN, T., KÄHÄRI, A. K., KEENAN, S., MARTIN, F. J., MAUREL, T., MCLAREN, W., MURPHY, D. N., NAG, R., OVERDUIN, B., PARKER, A., PATRICIO, M., PERRY, E., PIGNATELLI, M., RIAT, H. S., SHEPPARD, D., TAYLOR, K., THORMANN, A., VULLO, A., WILDER, S. P., ZADISSA, A., AKEN, B. L., BIRNEY, E., HARROW, J., KINSELLA, R., MUFFATO, M., RUFFIER, M., SEARLE, S. M. J., SPUDICH, G., TREVANION, S. J., YATES, A., ZERBINO, D. R. & FLICEK, P. 2015. Ensembl 2015. *Nucleic acids research,* 43**,** D662-D669.

CYBULSKY, A. V. & CYR, M. D. 1993. Phosphatidylcholine-directed phospholipase C: activation by complement C5b-9. *Am J Physiol,* 265**,** F551-60.

CYBULSKY, A. V., SALANT, D. J., QUIGG, R. J., BADALAMENTI, J. & BONVENTRE, J. V. 1989. Complement C5b-9 complex activates phospholipases in glomerular epithelial cells. *Am J Physiol,* 257**,** F826-36.

DAVIS, A. E., 3RD, HARRISON, R. A. & LACHMANN, P. J. 1984. Physiologic inactivation of fluid phase C3b: isolation and structural analysis of C3c, C3d,g (alpha 2D), and C3g. *J Immunol,* 132**,** 1960-6.

DAVYDOV, E. V., GOODE, D. L., SIROTA, M., COOPER, G. M., SIDOW, A. & BATZOGLOU, S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol,* 6**,** e1001025.

DECAFFMEYER, M., SHULGA, Y. V., DICU, A. O., THOMAS, A., TRUANT, R., TOPHAM, M. K., BRASSEUR, R. & EPAND, R. M. 2008. Determination of the topology of the hydrophobic segment of mammalian diacylglycerol kinase epsilon in a cell membrane and its relationship to predictions from modeling. *J Mol Biol,* 383**,** 797-809.

DELVAEYE, M., NORIS, M., DE VRIESE, A., ESMON, C. T., ESMON, N. L., FERRELL, G., DEL-FAVERO, J., PLAISANCE, S., CLAES, B., LAMBRECHTS, D., ZOJA, C., REMUZZI, G. & CONWAY, E. M. 2009. Thrombomodulin Mutations in Atypical Hemolytic–Uremic Syndrome. *New England Journal of Medicine,* 361**,** 345-357.

DEN DUNNEN, J. T. & ANTONARAKIS, S. E. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat,* 15**,** 7-12.

DEPRISTO, M. A., BANKS, E., POPLIN, R., GARIMELLA, K. V., MAGUIRE, J. R., HARTL, C., PHILIPPAKIS, A. A., DEL ANGEL, G., RIVAS, M. A., HANNA, M., MCKENNA, A., FENNELL, T. J., KERNYTSKY, A. M., SIVACHENKO, A. Y., CIBULSKIS, K., GABRIEL, S. B., ALTSHULER, D. & DALY, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet,* 43**,** 491-498.

DISCIPIO, R. G. 1992. Ultrastructures and interactions of complement factors H and I. *J Immunol,* 149**,** 2592-9.

DISCIPIO, R. G., GEHRING, M. R., PODACK, E. R., KAN, C. C., HUGLI, T. E. & FEY, G. H. 1984. Nucleotide sequence of cDNA and derived amino acid sequence of human complement component C9. *Proc Natl Acad Sci U S A,* 81**,** 7298-302.

DOLLEY, G., LAMARCHE, B., DESPRES, J. P., BOUCHARD, C., PERUSSE, L. & VOHL, M. C. 2009. Phosphoinositide cycle gene polymorphisms affect the plasma lipid profile in the Quebec Family Study. *Mol Genet Metab,* 97**,** 149-54.

DONG, C., WEI, P., JIAN, X., GIBBS, R., BOERWINKLE, E., WANG, K. & LIU, X. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet,* 24**,** 2125-37.

DRAGON-DUREY, M. A., BLANC, C., MARLIOT, F., LOIRAT, C., BLOUIN, J., SAUTES-FRIDMAN, C., FRIDMAN, W. H. & FREMEAUX-BACCHI, V. 2009. The high frequency of complement factor H related CFHR1 gene deletion is restricted to specific subgroups of patients with atypical haemolytic uraemic syndrome. *Journal of Medical Genetics,* 46**,** 447-50.

EBERMANN, I., WIESEN, M. H., ZRENNER, E., LOPEZ, I., PIGEON, R., KOHL, S., LOWENHEIM, H., KOENEKOOP, R. K. & BOLZ, H. J. 2009. GPR98 mutations cause Usher syndrome type 2 in males. *J Med Genet,* 46**,** 277-80.

EL-HUSSEINI, A., HANNAN, S., AWAD, A., JENNINGS, S., CORNEA, V. & SAWAYA, B. P. 2015. Thrombotic Microangiopathy in Systemic Lupus Erythematosus: Efficacy of Eculizumab. *American Journal of Kidney Diseases,* 65**,** 127-130.

ELDERING, E., SPEK, C. A., ABERSON, H. L., GRUMMELS, A., DERKS, I. A., DE VOS, A. F., MCELGUNN, C. J. & SCHOUTEN, J. P. 2003. Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. *Nucleic Acids Research,* 31**,** e153.

ELIMAM, H., PAPILLON, J., TAKANO, T. & CYBULSKY, A. V. 2013. Complement-mediated activation of calcium-independent phospholipase A2gamma: role of protein kinases and phosphorylation. *J Biol Chem,* 288**,** 3871-85.

EREMINA, V., JEFFERSON, J. A., KOWALEWSKA, J., HOCHSTER, H., HAAS, M., WEISSTUCH, J., RICHARDSON, C., KOPP, J. B., KABIR, M. G., BACKX, P. H., GERBER, H. P., FERRARA, N., BARISONI, L., ALPERS, C. E. & QUAGGIN, S. E. 2008. VEGF inhibition and renal thrombotic microangiopathy. *N Engl J Med,* 358**,** 1129-36.

ERLICH, Y., EDVARDSON, S., HODGES, E., ZENVIRT, S., THEKKAT, P., SHAAG, A., DOR, T., HANNON, G. J. & ELPELEG, O. 2011. Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. *Genome Research,* 21**,** 658-664.

ESAPA, C. T., BENSON, M. A., SCHRODER, J. E., MARTIN-RENDON, E., BROCKINGTON, M., BROWN, S. C., MUNTONI, F., KROGER, S. & BLAKE, D. J. 2002. Functional requirements for fukutin-related protein in the Golgi apparatus. *Hum Mol Genet,* 11**,** 3319-31.

ESPARZA-GORDILLO, J., GOICOECHEA DE JORGE, E., BUIL, A., CARRERAS BERGES, L., LOPEZ-TRASCASA, M., SANCHEZ-CORRAL, P. & RODRIGUEZ DE CORDOBA, S. 2005. Predisposition to atypical hemolytic uremic syndrome involves the concurrence of different susceptibility alleles in the regulators of complement activation gene cluster in 1q32. *Human Molecular Genetics,* 14**,** 703-12.

EYLER, S. J., MEYER, N. C., ZHANG, Y., XIAO, X., NESTER, C. M. & SMITH, R. J. 2013. A novel hybrid CFHR1/CFH gene causes atypical hemolytic uremic syndrome. *Pediatric Nephrology,* 28**,** 2221-5.

FAKHOURI, F., ROUMENINA, L., PROVOT, F., SALLÉE, M., CAILLARD, S., COUZI, L., ESSIG, M., RIBES, D., DRAGON-DUREY, M.-A., BRIDOUX, F., RONDEAU, E. & FRÉMEAUX-BACCHI, V. 2010. Pregnancy-Associated Hemolytic Uremic Syndrome Revisited in the Era of Complement Gene Mutations. *Journal of the American Society of Nephrology,* 21**,** 859-867.

FALCÃO, D. A., REIS, E. S., PAIXÃO-CAVALCANTE, D., AMANO, M. T., DELCOLLI, M. I. M. V., FLORIDO, M. P. C., ALBUQUERQUE, J. A. T., MORAES-VASCONCELOS, D., DUARTE, A. J., GRUMACH, A. S. & ISAAC, L. 2008. Deficiency of the Human Complement Regulatory Protein Factor H Associated with Low Levels of Component C9. *Scandinavian Journal of Immunology,* 68**,** 445-455.

FANG, H., WU, Y., NARZISI, G., O'RAWE, J. A., BARRON, L. T., ROSENBAUM, J., RONEMUS, M., IOSSIFOV, I., SCHATZ, M. C. & LYON, G. J. 2014. Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Med,* 6**,** 89.

FAUL, C., ASANUMA, K., YANAGIDA-ASANUMA, E., KIM, K. & MUNDEL, P. 2007. Actin up: regulation of podocyte structure and function by components of the actin cytoskeleton. *Trends Cell Biol,* 17**,** 428-37.

FEARON, D. T. 1978. Regulation by membrane sialic acid of beta1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway. *Proc Natl Acad Sci U S A,* 75**,** 1971-5.

FEARON, D. T., DAHA, M. R., WEILER, J. M. & AUSTEN, K. F. 1976. The natural modulation of the amplification phase of complement activation. *Transplant Rev,* 32**,** 12-25.

FENG, S., EYLER, S. J., ZHANG, Y., MAGA, T., NESTER, C. M., KROLL, M. H., SMITH, R. J. & AFSHAR-KHARGHAN, V. 2013. Partial ADAMTS13 deficiency in atypical hemolytic uremic syndrome. *Blood,* 122**,** 1487-93.

FERREIRA, V. P., HERBERT, A. P., CORTES, C., MCKEE, K. A., BLAUM, B. S., ESSWEIN, S. T., UHRIN, D., BARLOW, P. N., PANGBURN, M. K. & KAVANAGH, D. 2009. The binding of factor H to a complex of physiological polyanions and C3b on cells is impaired in atypical hemolytic uremic syndrome. *J Immunol,* 182**,** 7009-18.

FERREIRA, V. P., HERBERT, A. P., HOCKING, H. G., BARLOW, P. N. & PANGBURN, M. K. 2006. Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. *J Immunol,* 177**,** 6308-16.

FINN, R. D., BATEMAN, A., CLEMENTS, J., COGGILL, P., EBERHARDT, R. Y., EDDY, S. R., HEGER, A., HETHERINGTON, K., HOLM, L., MISTRY, J., SONNHAMMER, E. L. L., TATE, J. & PUNTA, M. 2014. Pfam: the protein families database. *Nucleic acids research,* 42**,** D222-D230.

FORNERIS, F., RICKLIN, D., WU, J., TZEKOU, A., WALLACE, R. S., LAMBRIS, J. D. & GROS, P. 2010. Structures of C3b in complex with factors B and D give insight into complement convertase formation. *Science,* 330**,** 1816-20.

FRANCIS, N. J., MCNICHOLAS, B., AWAN, A., WALDRON, M., REDDAN, D., SADLIER, D., KAVANAGH, D., STRAIN, L., MARCHBANK, K. J., HARRIS, C. L. & GOODSHIP, T. H. 2012. A novel hybrid CFH/CFHR3 gene generated by a microhomology-mediated deletion in familial atypical hemolytic uremic syndrome. *Blood,* 119**,** 591-601.

FREMEAUX-BACCHI, V., DRAGON-DUREY, M. A., BLOUIN, J., VIGNEAU, C., KUYPERS, D., BOUDAILLIEZ, B., LOIRAT, C., RONDEAU, E. & FRIDMAN, W. H. 2004. Complement factor I: a susceptibility gene for atypical haemolytic uraemic syndrome. *Journal of Medical Genetics,* 41**,** e84.

FREMEAUX-BACCHI, V., FAKHOURI, F., GARNIER, A., BIENAIMÉ, F., DRAGON-DUREY, M.-A., NGO, S., MOULIN, B., SERVAIS, A., PROVOT, F., ROSTAING, L., BURTEY, S., NIAUDET, P., DESCHÊNES, G., LEBRANCHU, Y., ZUBER, J. & LOIRAT, C. 2013. Genetics and Outcome of Atypical Hemolytic Uremic Syndrome: A Nationwide French Series Comparing Children and Adults. *Clinical Journal of the American Society of Nephrology,* 8**,** 554-562.

FREMEAUX-BACCHI, V., KEMP, E. J., GOODSHIP, J. A., DRAGON-DUREY, M. A., STRAIN, L., LOIRAT, C., DENG, H. W. & GOODSHIP, T. H. 2005. The development of atypical haemolytic-uraemic syndrome is influenced by susceptibility factors in factor H and membrane cofactor protein: evidence from two independent cohorts. *J Med Genet,* 42**,** 852-6.

FREMEAUX-BACCHI, V., MILLER, E. C., LISZEWSKI, M. K., STRAIN, L., BLOUIN, J., BROWN, A. L., MOGHAL, N., KAPLAN, B. S., WEISS, R. A., LHOTTA, K., KAPUR, G., MATTOO, T., NIVET, H., WONG, W., GIE, S., HURAULT DE LIGNY, B., FISCHBACH, M., GUPTA, R., HAUHART, R., MEUNIER, V., LOIRAT, C., DRAGON-DUREY, M. A., FRIDMAN, W. H., JANSSEN, B. J., GOODSHIP, T. H. & ATKINSON, J. P. 2008. Mutations in complement C3 predispose to development of atypical hemolytic uremic syndrome. *Blood,* 112**,** 4948-52.

FREMEAUX-BACCHI, V., MOULTON, E. A., KAVANAGH, D., DRAGON-DUREY, M. A., BLOUIN, J., CAUDY, A., ARZOUK, N., CLEPER, R., FRANCOIS,

M., GUEST, G., POURRAT, J., SELIGMAN, R., FRIDMAN, W. H., LOIRAT, C. & ATKINSON, J. P. 2006. Genetic and functional analyses of membrane cofactor protein (CD46) mutations in atypical hemolytic uremic syndrome. *Journal of the American Society of Nephrology,* 17**,** 2017-25.

FRIMAT, M., ROUMENINA, L. T., TABARIN, F., HALBWACHS-MECARELLI, L. & FREMEAUX-BACCHI, V. 2012. Membrane cofactor protein (MCP) haplotype, which predisposes to atypical hemolytic and uremic syndrome, has no consequence on neutrophils and endothelial cells MCP levels or on HUVECs ability to activate complement. *Immunobiology,* 217**,** 1187-1188.

FUCHS, A., ATKINSON, J. P., FREMEAUX-BACCHI, V. & KEMPER, C. 2009. CD46-induced human Treg enhance B-cell responses. *Eur J Immunol,* 39**,** 3097-109.

FULLER, C. W., MIDDENDORF, L. R., BENNER, S. A., CHURCH, G. M., HARRIS, T., HUANG, X., JOVANOVICH, S. B., NELSON, J. R., SCHLOSS, J. A., SCHWARTZ, D. C. & VEZENOV, D. V. 2009. The challenges of sequencing by synthesis. *Nat Biotech,* 27**,** 1013-1023.

FURLAN, M., ROBLES, R. & LAMMLE, B. 1996. Partial purification and characterization of a protease from human plasma cleaving von Willebrand factor to fragments produced by in vivo proteolysis. *Blood,* 87**,** 4223-34.

FURUKAWA, T., KUBOKI, Y., TANJI, E., YOSHIDA, S., HATORI, T., YAMAMOTO, M., SHIBATA, N., SHIMIZU, K., KAMATANI, N. & SHIRATORI, K. 2011. Whole-exome sequencing uncovers frequent GNAS mutations in intraductal papillary mucinous neoplasms of the pancreas. *Scientific Reports,* 1.

GALE, A. J. & GRIFFIN, J. H. 2004. Characterization of a thrombomodulin binding site on protein C and its comparison to an activated protein C binding site for factor Va. *Proteins,* 54**,** 433-41.

GARRISON, E. & MARTH, G. 2012. Haplotype-based variant detection from short-read sequencing.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S. E., WILKINS, M., APPEL, R. & BAIROCH, A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. *In:* WALKER, J. (ed.) *The Proteomics Protocols Handbook.* Humana Press.

GBADEGESIN, R. A., LAVIN, P. J., HALL, G., BARTKOWIAK, B., HOMSTAD, A., JIANG, R., WU, G., BYRD, A., LYNN, K., WOLFISH, N., OTTATI, C., STEVENS, P., HOWELL, D., CONLON, P. & WINN, M. P. 2012. Inverted formin 2 mutations with variable expression in patients with sporadic and hereditary focal and segmental glomerulosclerosis. *Kidney International,* 81**,** 94-9.

GE-HEALTHCARE. 2013a. *Instructions 52-1308-00 BB: PD-10 Desalting Columns* [Online]. Available: http://www.gelifesciences.com/webapp/wcs/stores/servlet/productById/en/GELifeSciences/17085101 [Accessed 15.03.13.

GE-HEALTHCARE. 2013b. *Instructions 71-7006-00 AW: HiTrap™ NHS-activated HP* [Online]. Available:

http://www.gelifesciences.com/webapp/wcs/stores/servlet/productById/en/GELifeSciences/17071601 [Accessed 26.02.13.

GE-HEALTHCARE. 2015. Antibody Purification Handbook. Available: https://http://www.gelifesciences.com/gehcls_images/GELS/Related Content/Files/1335538533725/litdoc18103746_20150210215432.pdf [Accessed 24.03.15].

GILISSEN, C., HOISCHEN, A., BRUNNER, H. G. & VELTMAN, J. A. 2012. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet,* 20**,** 490-7.

GOICOECHEA DE JORGE, E., CAESAR, J. J., MALIK, T. H., PATEL, M., COLLEDGE, M., JOHNSON, S., HAKOBYAN, S., MORGAN, B. P., HARRIS, C. L., PICKERING, M. C. & LEA, S. M. 2013. Dimerization of complement factor H-related proteins modulates complement activation in vivo. *Proc Natl Acad Sci U S A,* 110**,** 4685-90.

GOICOECHEA DE JORGE, E., HARRIS, C. L., ESPARZA-GORDILLO, J., CARRERAS, L., ARRANZ, E. A., GARRIDO, C. A., LOPEZ-TRASCASA, M., SANCHEZ-CORRAL, P., MORGAN, B. P. & RODRIGUEZ DE CORDOBA, S. 2007. Gain-of-function mutations in complement factor B are associated with atypical hemolytic uremic syndrome. *Proc Natl Acad Sci U S A,* 104**,** 240-5.

GOICOECHEA DE JORGE, E., MACOR, P., PAIXÃO-CAVALCANTE, D., ROSE, K. L., TEDESCO, F., COOK, H. T., BOTTO, M. & PICKERING, M. C. 2011. The Development of Atypical Hemolytic Uremic Syndrome Depends on Complement C5. *Journal of the American Society of Nephrology,* 22**,** 137-145.

GOLDBERGER, G., BRUNS, G. A., RITS, M., EDGE, M. D. & KWIATKOWSKI, D. J. 1987. Human complement factor I: analysis of cDNA-derived primary structure and assignment of its gene to chromosome 4. *J Biol Chem,* 262**,** 10065-71.

GOLDSTEIN, D. B., ALLEN, A., KEEBLER, J., MARGULIES, E. H., PETROU, S., PETROVSKI, S. & SUNYAEV, S. 2013. Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet,* 14**,** 460-70.

GOODE, B. L. & ECK, M. J. 2007. Mechanism and Function of Formins in the Control of Actin Assembly. *Annual Review of Biochemistry,* 76**,** 593-627.

GORDON, D. L., KAUFMAN, R. M., BLACKMORE, T. K., KWONG, J. & LUBLIN, D. M. 1995. Identification of complement regulatory domains in human factor H. *J Immunol,* 155**,** 348-56.

GOULD, C. J., MAITI, S., MICHELOT, A., GRAZIANO, B. R., BLANCHOIN, L. & GOODE, B. L. 2011. The formin DAD domain plays dual roles in autoinhibition and actin nucleation. *Curr Biol,* 21**,** 384-90.

GRAY, V. E., KUKURBA, K. R. & KUMAR, S. 2012. Performance of computational tools in evaluating the functional impact of laboratory-induced amino acid mutations. *Bioinformatics,* 28**,** 2093-2096.

GREEN, R. 2010. Ins and outs of cellular cobalamin transport. *Blood,* 115**,** 1476-1477.

GUO, J., XU, N., LI, Z., ZHANG, S., WU, J., KIM, D. H., SANO MARMA, M., MENG, Q., CAO, H., LI, X., SHI, S., YU, L., KALACHIKOV, S., RUSSO, J. J., TURRO, N. J. & JU, J. 2008. Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc Natl Acad Sci U S A,* 105**,** 9145-50.

GUREL, P. S., GE, P., GRINTSEVICH, E. E., SHU, R., BLANCHOIN, L., ZHOU, Z. H., REISLER, E. & HIGGS, H. N. 2014. INF2-mediated severing through actin filament encirclement and disruption. *Curr Biol,* 24**,** 156-64.

HADDERS, M. A., BUBECK, D., ROVERSI, P., HAKOBYAN, S., FORNERIS, F., MORGAN, B. P., PANGBURN, M. K., LLORCA, O., LEA, S. M. & GROS, P. 2012. Assembly and Regulation of the Membrane Attack Complex Based on Structures of C5b6 and sC5b9. *Cell Rep,* 1**,** 200-7.

HAGEMAN, G. S., ANDERSON, D. H., JOHNSON, L. V., HANCOX, L. S., TAIBER, A. J., HARDISTY, L. I., HAGEMAN, J. L., STOCKMAN, H. A., BORCHARDT, J. D., GEHRS, K. M., SMITH, R. J., SILVESTRI, G., RUSSELL, S. R., KLAVER, C. C., BARBAZETTO, I., CHANG, S., YANNUZZI, L. A., BARILE, G. R., MERRIAM, J. C., SMITH, R. T., OLSH, A. K., BERGERON, J., ZERNANT, J., MERRIAM, J. E., GOLD, B., DEAN, M. & ALLIKMETS, R. 2005. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci U S A,* 102**,** 7227-32.

HALPERIN, J. A., TARATUSKA, A. & NICHOLSON-WELLER, A. 1993. Terminal complement complex C5b-9 stimulates mitogenesis in 3T3 cells. *J Clin Invest,* 91**,** 1974-8.

HANSCH, G. M., GEMSA, D. & RESCH, K. 1985. Induction of prostanoid synthesis in human platelets by the late complement components C5b-9 and channel forming antibiotic nystatin: inhibition of the reacylation of liberated arachidonic acid. *J Immunol,* 135**,** 1320-4.

HANSCH, G. M., SEITZ, M., MARTINOTTI, G., BETZ, M., RAUTERBERG, E. W. & GEMSA, D. 1984. Macrophages release arachidonic acid, prostaglandin E2, and thromboxane in response to late complement components. *J Immunol,* 133**,** 2145-50.

HARRIS, C. L. 2000. Functional assays for complement regulators. *Methods Mol Biol,* 150**,** 83-101.

HART, T. C., PRICE, J. A., BOBBY, P. L., PETTENATI, M. J., SHASHI, V., VON KAP HERR, C. & VAN DYKE, T. E. 1999. Cytogenetic assignment and physical mapping of the human DGKE gene to chromosome 17q22. *Genomics,* 56**,** 233-5.

HASMATS, J., GRÉEN, H., OREAR, C., VALIDIRE, P., HUSS, M., KÄLLER, M. & LUNDEBERG, J. 2014. Assessment of Whole Genome Amplification for Sequence Capture and Massively Parallel Sequencing. *PLoS One,* 9**,** e84785.

HATTORI, R., HAMILTON, K. K., MCEVER, R. P. & SIMS, P. J. 1989. Complement proteins C5b-9 induce secretion of high molecular weight multimers of endothelial von Willebrand factor and translocation of granule membrane protein GMP-140 to the cell surface. *J Biol Chem,* 264**,** 9053-60.

HEERINGA, S. F., MÖLLER, C. C., DU, J., YUE, L., HINKES, B., CHERNIN, G., VLANGOS, C. N., HOYER, P. F., REISER, J. & HILDEBRANDT, F. 2009. A Novel TRPC6 Mutation That Causes Childhood FSGS. *PLoS One,* 4**,** e7771.

HEINEN, S., SANCHEZ-CORRAL, P., JACKSON, M. S., STRAIN, L., GOODSHIP, J. A., KEMP, E. J., SKERKA, C., JOKIRANTA, T. S., MEYERS, K., WAGNER, E., ROBITAILLE, P., ESPARZA-GORDILLO, J., RODRIGUEZ DE CORDOBA, S., ZIPFEL, P. F. & GOODSHIP, T. H. J. 2006. De novo gene conversion in the RCA gene cluster (1q32) causes mutations in complement factor H associated with atypical hemolytic uremic syndrome. *Human Mutation,* 27**,** 292-293.

HELLWAGE, J., JOKIRANTA, T. S., KOISTINEN, V., VAARALA, O., MERI, S. & ZIPFEL, P. F. 1999. Functional properties of complement factor H-related proteins FHR-3 and FHR-4: binding to the C3d region of C3b and differential regulation by heparin. *FEBS Lett,* 462**,** 345-52.

HEURICH, M., MARTINEZ-BARRICARTE, R., FRANCIS, N. J., ROBERTS, D. L., RODRIGUEZ DE CORDOBA, S., MORGAN, B. P. & HARRIS, C. L. 2011. Common polymorphisms in C3, factor B, and factor H collaborate to determine systemic complement activity and disease risk. *Proc Natl Acad Sci U S A,* 108**,** 8761-6.

HEWITT, J. E. 2009. Abnormal glycosylation of dystroglycan in human genetic disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* 1792**,** 853-861.

HILLMEN, P., YOUNG, N. S., SCHUBERT, J., BRODSKY, R. A., SOCIÉ, G., MUUS, P., RÖTH, A., SZER, J., ELEBUTE, M. O., NAKAMURA, R., BROWNE, P., RISITANO, A. M., HILL, A., SCHREZENMEIER, H., FU, C.-L., MACIEJEWSKI, J., ROLLINS, S. A., MOJCIK, C. F., ROTHER, R. P. & LUZZATTO, L. 2006. The Complement Inhibitor Eculizumab in Paroxysmal Nocturnal Hemoglobinuria. *New England Journal of Medicine,* 355**,** 1233-1243.

HOCHSMANN, B., DOHNA-SCHWAKE, C., KYRIELEIS, H. A., PANNICKE, U. & SCHREZENMEIER, H. 2014. Targeted therapy with eculizumab for inherited CD59 deficiency. *N Engl J Med,* 370**,** 90-2.

HOCKING, H. G., HERBERT, A. P., KAVANAGH, D., SOARES, D. C., FERREIRA, V. P., PANGBURN, M. K., UHRIN, D. & BARLOW, P. N. 2008. Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations. *J Biol Chem,* 283**,** 9475-87.

HODGES, E., XUAN, Z., BALIJA, V., KRAMER, M., MOLLA, M. N., SMITH, S. W., MIDDLE, C. M., RODESCH, M. J., ALBERT, T. J., HANNON, G. J. & MCCOMBIE, W. R. 2007. Genome-wide in situ exon capture for selective resequencing. *Nat Genet,* 39**,** 1522-7.

HOFMANN, T., OBUKHOV, A. G., SCHAEFER, M., HARTENECK, C., GUDERMANN, T. & SCHULTZ, G. 1999. Direct activation of human TRPC6 and TRPC3 channels by diacylglycerol. *Nature,* 397**,** 259-63.

HOLERS, V. M. 2014. Complement and its receptors: new insights into human disease. *Annu Rev Immunol,* 32**,** 433-59.

HOLLEGAARD, M. V., GRAUHOLM, J., NIELSEN, R., GROVE, J., MANDRUP, S. & HOUGAARD, D. M. 2013. Archived neonatal dried blood spot samples can be used for accurate whole genome and exome-targeted next-generation sequencing. *Molecular Genetics and Metabolism,* 110**,** 65-72.

HOLLIDAY, R. 2007. A mechanism for gene conversion in fungi. *Genetics Research,* 89**,** 285-307.

HORIUCHI, T., NISHIZAKA, H., KOJIMA, T., SAWABE, T., NIHO, Y., SCHNEIDER, P. M., INABA, S., SAKAI, K., HAYASHI, K., HASHIMURA, C. & FUKUMORI, Y. 1998. A non-sense mutation at Arg95 is predominant in complement 9 deficiency in Japanese. *J Immunol,* 160**,** 1509-13.

HOURCADE, D., HOLERS, V. M. AND ATKINSON, J. P. 1989. The Regulators of Complement Activation (RCA) Gene Cluster. *In:* DIXON, F. J. (ed.) *Advances in Immunology.* San Diego, CA, USA.: ACADEMIC PRESS, INC.

HOURCADE, D. E., MITCHELL, L., KUTTNER-KONDO, L. A., ATKINSON, J. P. & MEDOF, M. E. 2002. Decay-accelerating factor (DAF), complement receptor 1 (CR1), and factor H dissociate the complement AP C3 convertase (C3bBb) via sites on the type A domain of Bb. *J Biol Chem,* 277**,** 1107-12.

HUANG, Y., QIAO, F., ABAGYAN, R., HAZARD, S. & TOMLINSON, S. 2006. Defining the CD59-C9 Binding Interaction. *Journal of Biological Chemistry,* 281**,** 27398-27404.

ISAAC, L. & ISENMAN, D. E. 1992. Structural requirements for thioester bond formation in human complement component C3. Reassessment of the role of thioester bond integrity on the conformation of C3. *J Biol Chem,* 267**,** 10062-9.

JACKMAN, R. W., BEELER, D. L., FRITZE, L., SOFF, G. & ROSENBERG, R. D. 1987. Human thrombomodulin gene is intron depleted: nucleic acid sequences of the cDNA and gene predict protein structure and suggest sites of regulatory control. *Proc Natl Acad Sci U S A,* 84**,** 6425-9.

JANSSEN, B. J., CHRISTODOULIDOU, A., MCCARTHY, A., LAMBRIS, J. D. & GROS, P. 2006. Structure of C3b reveals conformational changes that underlie complement activity. *Nature,* 444**,** 213-6.

JANSSEN, B. J., HUIZINGA, E. G., RAAIJMAKERS, H. C., ROOS, A., DAHA, M. R., NILSSON-EKDAHL, K., NILSSON, B. & GROS, P. 2005. Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature,* 437**,** 505-11.

JOZSI, M., LICHT, C., STROBEL, S., ZIPFEL, S. L., RICHTER, H., HEINEN, S., ZIPFEL, P. F. & SKERKA, C. 2008. Factor H autoantibodies in atypical hemolytic uremic syndrome correlate with CFHR1/CFHR3 deficiency. *Blood,* 111**,** 1512-4.

JOZSI, M. & ZIPFEL, P. F. 2008. Factor H family proteins and human diseases. *Trends Immunol,* 29**,** 380-7.

JU, J., KIM, D. H., BI, L., MENG, Q., BAI, X., LI, Z., LI, X., MARMA, M. S., SHI, S., WU, J., EDWARDS, J. R., ROMU, A. & TURRO, N. J. 2006. Four-color DNA

sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc Natl Acad Sci U S A,* 103**,** 19635-40.

KAMAR, N., RISCHMANN, P., GUILBEAU-FRUGIER, C., SALLUSTO, F., KHEDIS, M., DELISLE, M.-B., NOURY, D., FORT, M. & ROSTAING, L. 2008. Successful Retransplantation of a Kidney Allograft Affected by Thrombotic Microangiopathy Into a Second Transplant Recipient. *American Journal of Kidney Diseases,* 52**,** 591-594.

KAPLAN, B. S., TROMPETER, R. S. & MOAKE, J. L. 1992. *Hemolytic Uremic Syndrome and Thrombotic Thrombocytopenic Purpura,* NY, USA, Marcel Dekker, Inc.

KAPLAN, J. M., H KIM, S., NORTH, K. N., RENNKE, H., A CORREIA, L., TONG, H.-Q., MATHIS, B. J., RODRIGUEZ-PEREZ, J.-C., ALLEN, P. G., BEGGS, A. H. & POLLAK, M. R. 2000. Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal segmental glomerulosclerosis. *Nat Genet,* 24**,** 251-256.

KAROLCHIK, D., BARBER, G. P., CASPER, J., CLAWSON, H., CLINE, M. S., DIEKHANS, M., DRESZER, T. R., FUJITA, P. A., GURUVADOO, L., HAEUSSLER, M., HARTE, R. A., HEITNER, S., HINRICHS, A. S., LEARNED, K., LEE, B. T., LI, C. H., RANEY, B. J., RHEAD, B., ROSENBLOOM, K. R., SLOAN, C. A., SPEIR, M. L., ZWEIG, A. S., HAUSSLER, D., KUHN, R. M. & KENT, W. J. 2014. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Research,* 42**,** D764-D770.

KASS, E. M. & JASIN, M. 2010. Collaboration and competition between DNA double-strand break repair pathways. *FEBS Letters,* 584**,** 3703-3708.

KAVANAGH, D. & GOODSHIP, T. 2010. Genetics and complement in atypical HUS. *Pediatric Nephrology,* 25**,** 2431-2442.

KAVANAGH, D., GOODSHIP, T. H. & RICHARDS, A. 2013. Atypical Hemolytic Uremic Syndrome. *Semin Nephrol,* 33**,** 508-530.

KAVANAGH, D., KEMP, E. J., MAYLAND, E., WINNEY, R. J., DUFFIELD, J. S., WARWICK, G., RICHARDS, A., WARD, R., GOODSHIP, J. A. & GOODSHIP, T. H. 2005. Mutations in complement factor I predispose to development of atypical hemolytic uremic syndrome. *Journal of the American Society of Nephrology,* 16**,** 2150-5.

KAVANAGH, D., RAMAN, S. & SHEERIN, N. S. 2014. Management of hemolytic uremic syndrome. *F1000Prime Rep,* 6**,** 119.

KAVANAGH, D., RICHARDS, A., NORIS, M., HAUHART, R., LISZEWSKI, M. K., KARPMAN, D., GOODSHIP, J. A., FREMEAUX-BACCHI, V., REMUZZI, G., GOODSHIP, T. H. J. & ATKINSON, J. P. 2008. Characterization of mutations in complement factor I (CFI) associated with hemolytic uremic syndrome. *Molecular Immunology,* 45**,** 95-105.

KELLEY, L. A. & STERNBERG, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* 4**,** 363-71.

KELLY, U., YU, L., KUMAR, P., DING, J.-D., JIANG, H., HAGEMAN, G. S., ARSHAVSKY, V. Y., FRANK, M. M., HAUSER, M. A. & RICKMAN, C. B. 2010. Heparan Sulfate, Including that in Bruch's Membrane, Inhibits the Complement

Alternative Pathway: Implications for Age-related Macular Degeneration. *Journal of immunology (Baltimore, Md. : 1950),* 185**,** 5486-5494.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Research,* 12**,** 996-1006.

KENTOUCHE, K., BUDDE, U., FURLAN, M., SCHARFE, V., SCHNEPPENHEIM, R. & ZINTL, F. 2002. Remission of thrombotic thrombocytopenic purpura in a patient with compound heterozygous deficiency of von Willebrand factor-cleaving protease by infusion of solvent/detergent plasma. *Acta Paediatr,* 91**,** 1056-9.

KEREM, B., ROMMENS, J. M., BUCHANAN, J. A., MARKIEWICZ, D., COX, T. K., CHAKRAVARTI, A., BUCHWALD, M. & TSUI, L. C. 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science,* 245**,** 1073-80.

KESSNER, D., CHAMBERS, M., BURKE, R., AGUS, D. & MALLICK, P. 2008. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics,* 24**,** 2534-6.

KHALIL, R. A. 2013. Protein Kinase C Inhibitors as Modulators of Vascular Function and their Application in Vascular Disease. *Pharmaceuticals (Basel),* 6**,** 407-39.

KIKKAWA, U., KISHIMOTO, A. & NISHIZUKA, Y. 1989. The protein kinase C family: heterogeneity and its implications. *Annu Rev Biochem,* 58**,** 31-44.

KILGORE, K. S., SCHMID, E., SHANLEY, T. P., FLORY, C. M., MAHESWARI, V., TRAMONTINI, N. L., COHEN, H., WARD, P. A., FRIEDL, H. P. & WARREN, J. S. 1997. Sublytic concentrations of the membrane attack complex of complement induce endothelial interleukin-8 and monocyte chemoattractant protein-1 through nuclear factor-kappa B activation. *Am J Pathol,* 150**,** 2019-31.

KITAGAWA, H. & PAULSON, J. C. 1994. Differential expression of five sialyltransferase genes in human tissues. *Journal of Biological Chemistry,* 269**,** 17872-17878.

KLEIN, R. J., ZEISS, C., CHEW, E. Y., TSAI, J.-Y., SACKLER, R. S., HAYNES, C., HENNING, A. K., SANGIOVANNI, J. P., MANE, S. M., MAYNE, S. T., BRACKEN, M. B., FERRIS, F. L., OTT, J., BARNSTABLE, C. & HOH, J. 2005. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science (New York, N.Y.),* 308**,** 385-389.

KOSTER, F., LEVIN, J., WALKER, L., TUNG, K. S., GILMAN, R. H., RAHAMAN, M. M., MAJID, M. A., ISLAM, S. & WILLIAMS, R. C., JR. 1978. Hemolytic-uremic syndrome after shigellosis. Relation to endotoxemia and circulating immune complexes. *N Engl J Med,* 298**,** 927-33.

KOVAR, D. R., HARRIS, E. S., MAHAFFY, R., HIGGS, H. N. & POLLARD, T. D. 2006. Control of the Assembly of ATP- and ADP-Actin by Formins and Profilin. *Cell,* 124**,** 423-435.

KOZAREWA, I., NING, Z., QUAIL, M. A., SANDERS, M. J., BERRIMAN, M. & TURNER, D. J. 2009. Amplification-free Illumina sequencing-library preparation

facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Meth,* 6**,** 291-295.

KRAUS, S. & FISHELSON, Z. 2000. Cell desensitization by sublytic C5b-9 complexes and calcium ionophores depends on activation of protein kinase C. *Eur J Immunol,* 30**,** 1272-80.

KRAWITZ, P., RODELSPERGER, C., JAGER, M., JOSTINS, L., BAUER, S. & ROBINSON, P. N. 2010. Microindel detection in short-read sequence data. *Bioinformatics,* 26**,** 722-9.

KU, C. S., COOPER, D. N., POLYCHRONAKOS, C., NAIDOO, N., WU, M. C. & SOONG, R. 2012. Exome sequencing: Dual role as a discovery and diagnostic tool. *Annals of Neurology,* 71**,** 5-14.

KUHN, S., SKERKA, C. & ZIPFEL, P. F. 1995. Mapping of the complement regulatory domains in the human factor H-like protein 1 and in factor H1. *J Immunol,* 155**,** 5663-70.

KUHN, S. & ZIPFEL, P. F. 1996. Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. *Eur J Immunol,* 26**,** 2383-7.

LAMBRIS, J. D., AVILA, D., BECHERER, J. D. & MULLER-EBERHARD, H. J. 1988. A discontinuous factor H binding site in the third component of complement as delineated by synthetic peptides. *J Biol Chem,* 263**,** 12147-50.

LANDER, E. S. & BOTSTEIN, D. 1987. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science,* 236**,** 1567-70.

LAURIN, L. P., LU, M., MOTTL, A. K., BLYTH, E. R., POULTON, C. J. & WECK, K. E. 2014. Podocyte-associated gene mutation screening in a heterogeneous cohort of patients with sporadic focal segmental glomerulosclerosis. *Nephrol Dial Transplant*.

LEE, H. K., HAN, K. H., JUNG, Y. H., KANG, H. G., MOON, K. C., HA, I. S., CHOI, Y. & CHEONG, H. I. 2011. Variable renal phenotype in a family with an INF2 mutation. *Pediatric Nephrology,* 26**,** 73-6.

LEMAIRE, M. 2015. *Of sugars & fats. Unbiased genomic studies suggest novel mechanism(s) causing atypical HUS. Nephrology city-wide rounds.* [Online]. Available: http://mediasite.otn.ca/Mediasite/Play/c2ab6cd50d8143f993a7a0d5a9c4fe531d?catalog=fd668812-d87c-47f9-b1ba-6d979fed9af4 [Accessed 19.04.15.

LEMAIRE, M., FREMEAUX-BACCHI, V., SCHAEFER, F., CHOI, M., TANG, W. H., LE QUINTREC, M., FAKHOURI, F., TAQUE, S., NOBILI, F., MARTINEZ, F., JI, W., OVERTON, J. D., MANE, S. M., NURNBERG, G., ALTMULLER, J., THIELE, H., MORIN, D., DESCHENES, G., BAUDOUIN, V., LLANAS, B., COLLARD, L., MAJID, M. A., SIMKOVA, E., NURNBERG, P., RIOUX-LECLERC, N., MOECKEL, G. W., GUBLER, M. C., HWA, J., LOIRAT, C. & LIFTON, R. P. 2013. Recessive mutations in DGKE cause atypical hemolytic-uremic syndrome. *Nature Genetics,* 45**,** 531-6.

LEONG, I. U., STUCKEY, A., LAI, D., SKINNER, J. R. & LOVE, D. R. 2015. Assessment of the predictive accuracy of five in silico prediction tools, alone or in

combination, and two metaservers to classify long QT syndrome gene mutations. *BMC Med Genet,* 16**,** 34.

LERNER-ELLIS, J. P., TIRONE, J. C., PAWELEK, P. D., DORE, C., ATKINSON, J. L., WATKINS, D., MOREL, C. F., FUJIWARA, T. M., MORAS, E., HOSACK, A. R., DUNBAR, G. V., ANTONICKA, H., FORGETTA, V., DOBSON, C. M., LECLERC, D., GRAVEL, R. A., SHOUBRIDGE, E. A., COULTON, J. W., LEPAGE, P., ROMMENS, J. M., MORGAN, K. & ROSENBLATT, D. S. 2006. Identification of the gene responsible for methylmalonic aciduria and homocystinuria, cblC type. *Nat Genet,* 38**,** 93-100.

LEVY, G. G., NICHOLS, W. C., LIAN, E. C., FOROUD, T., MCCLINTICK, J. N., MCGEE, B. M., YANG, A. Y., SIEMIENIAK, D. R., STARK, K. R., GRUPPO, R., SARODE, R., SHURIN, S. B., CHANDRASEKARAN, V., STABLER, S. P., SABIO, H., BOUHASSIRA, E. E., UPSHAW, J. D., JR., GINSBURG, D. & TSAI, H. M. 2001. Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura. *Nature,* 413**,** 488-94.

LI, H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics,* 30**,** 2843-2851.

LI, H. & DURBIN, R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics,* 26**,** 589-595.

LI, Q., LIU, X., GIBBS, R. A., BOERWINKLE, E., POLYCHRONAKOS, C. & QU, H.-Q. 2014. Gene-Specific Function Prediction for Non-Synonymous Mutations in Monogenic Diabetes Genes. *PLoS One,* 9**,** e104452.

LIFE-TECHNOLOGIES. 2009. XCell II™ Blot Module Available: https://tools.lifetechnologies.com/content/sfs/manuals/blotmod_pro.pdf [Accessed 24.03.15].

LIFE-TECHNOLOGIES. 2012. XCell SureLock® Mini-Cell Available: https://tools.lifetechnologies.com/content/sfs/manuals/surelock_man.pdf [Accessed 24.03.15].

LIPSKA, B. S., IATROPOULOS, P., MARANTA, R., CARIDI, G., OZALTIN, F., ANARAT, A., BALAT, A., GELLERMANN, J., TRAUTMANN, A., ERDOGAN, O., SAEED, B., EMRE, S., BOGDANOVIC, R., AZOCAR, M., BALASZ-CHMIELEWSKA, I., BENETTI, E., CALISKAN, S., MIR, S., MELK, A., ERTAN, P., BASKIN, E., JARDIM, H., DAVITAIA, T., WASILEWSKA, A., DROZDZ, D., SZCZEPANSKA, M., JANKAUSKIENE, A., HIGUITA, L. M., ARDISSINO, G., OZKAYA, O., KUZMA-MROCZKOWSKA, E., SOYLEMEZOGLU, O., RANCHIN, B., MEDYNSKA, A., TKACZYK, M., PECO-ANTIC, A., AKIL, I., JARMOLINSKI, T., FIRSZT-ADAMCZYK, A., DUSEK, J., SIMONETTI, G. D., GOK, F., GHEISSARI, A., EMMA, F., KRMAR, R. T., FISCHBACH, M., PRINTZA, N., SIMKOVA, E., MELE, C., GHIGGERI, G. M. & SCHAEFER, F. 2013. Genetic screening in adolescents with steroid-resistant nephrotic syndrome. *Kidney International,* 84**,** 206-13.

LISZEWSKI, M. K., LEUNG, M., CUI, W., SUBRAMANIAN, V. B., PARKINSON, J., BARLOW, P. N., MANCHESTER, M. & ATKINSON, J. P. 2000. Dissecting sites

important for complement regulatory activity in membrane cofactor protein (MCP; CD46). *J Biol Chem,* 275**,** 37692-701.

LISZEWSKI, M. K., LEUNG, M. K., SCHRAML, B., GOODSHIP, T. H. J. & ATKINSON, J. P. 2007. Modeling how CD46 deficiency predisposes to atypical hemolytic uremic syndrome. *Molecular Immunology,* 44**,** 1559-1568.

LISZEWSKI, M. K., POST, T. W. & ATKINSON, J. P. 1991. Membrane cofactor protein (MCP or CD46): newest member of the regulators of complement activation gene cluster. *Annu Rev Immunol,* 9**,** 431-55.

LIU, Y. & WEST, S. C. 2004. Happy Hollidays: 40th anniversary of the Holliday junction. *Nat Rev Mol Cell Biol,* 5**,** 937-944.

LODISH, H., BERK, A., ZIPURSKY, S. L., MATSUDAIRA, P., BALTIMORE, D. & DARNELL, J. 2000. The Dynamics of Actin Assembly. *Molecular Cell Biology.* 4th Edition ed. New York, USA: W. H. Freeman & CO.

LOIRAT, C. & FREMEAUX-BACCHI, V. 2011. Atypical hemolytic uremic syndrome. *Orphanet J Rare Dis,* 6**,** 60.

LOIRAT, C., NORIS, M. & FREMEAUX-BACCHI, V. 2008. Complement and the atypical hemolytic uremic syndrome in children. *Pediatric Nephrology,* 23**,** 1957-72.

LOIRAT, C., SALAND, J. & BITZAN, M. 2012. Management of hemolytic uremic syndrome. *La Presse Médicale,* 41**,** e115-e135.

LOVETT, D. H., HAENSCH, G. M., GOPPELT, M., RESCH, K. & GEMSA, D. 1987. Activation of glomerular mesangial cells by the terminal membrane attack complex of complement. *The Journal of Immunology,* 138**,** 2473-80.

LUBLIN, D. M., LISZEWSKI, M. K., POST, T. W., ARCE, M. A., LE BEAU, M. M., REBENTISCH, M. B., LEMONS, L. S., SEYA, T. & ATKINSON, J. P. 1988. Molecular cloning and chromosomal localization of human membrane cofactor protein (MCP). Evidence for inclusion in the multigene family of complement-regulatory proteins. *J Exp Med,* 168**,** 181-94.

LUECK, K., WASMUTH, S., WILLIAMS, J., HUGHES, T. R., MORGAN, B. P., LOMMATZSCH, A., GREENWOOD, J., MOSS, S. E. & PAULEIKHOFF, D. 2011. Sub-lytic C5b-9 induces functional changes in retinal pigment epithelial cells consistent with age-related macular degeneration. *Eye,* 25**,** 1074-1082.

MA, J.-L., KIM, E. M., HABER, J. E. & LEE, S. E. 2003. Yeast Mre11 and Rad1 Proteins Define a Ku-Independent Mechanism To Repair Double-Strand Breaks Lacking Overlapping End Sequences. *Molecular and Cellular Biology,* 23**,** 8820-8828.

MADEMAN, I., DECONINCK, T., DINOPOULOS, A., VOIT, T., SCHARA, U., DEVRIENDT, K., MEIJERS, B., LERUT, E., DE JONGHE, P. & BAETS, J. 2013. De novo INF2 mutations expand the genetic spectrum of hereditary neuropathy with glomerulopathy. *Neurology,* 81**,** 1953-8.

MADRID, R., ARANDA, J. F., RODRIGUEZ-FRATICELLI, A. E., VENTIMIGLIA, L., ANDRES-DELGADO, L., SHEHATA, M., FANAYAN, S., SHAHHEYDARI, H., GOMEZ, S., JIMENEZ, A., MARTIN-BELMONTE, F., BYRNE, J. A. & ALONSO,

M. A. 2010. The formin INF2 regulates basolateral-to-apical transcytosis and lumen formation in association with Cdc42 and MAL2. *Dev Cell,* 18**,** 814-27.

MAGA, T. K., MEYER, N. C., BELSHA, C., NISHIMURA, C. J., ZHANG, Y. & SMITH, R. J. H. 2011. A novel deletion in the RCA gene cluster causes atypical hemolytic uremic syndrome. *Nephrology Dialysis Transplantation,* 26**,** 739-741.

MAGA, T. K., NISHIMURA, C. J., WEAVER, A. E., FREES, K. L. & SMITH, R. J. 2010. Mutations in alternative pathway complement proteins in American patients with atypical hemolytic uremic syndrome. *Human Mutation,* 31**,** E1445-60.

MAGLOTT, D. R., FELDBLYUM, T. V., DURKIN, A. S. & NIERMAN, W. C. 1996. Radiation hybrid mapping of SNAP, PCSK2, and THBD (human chromosome 20p). *Mamm Genome,* 7**,** 400-1.

MAMANOVA, L., COFFEY, A. J., SCOTT, C. E., KOZAREWA, I., TURNER, E. H., KUMAR, A., HOWARD, E., SHENDURE, J. & TURNER, D. J. 2010. Target-enrichment strategies for next-generation sequencing. *Nat Methods,* 7**,** 111-8.

MANENTI, L., GNAPPI, E., VAGLIO, A., ALLEGRI, L., NORIS, M., BRESIN, E., PILATO, F. P., VALOTI, E., PASQUALI, S. & BUZIO, C. 2013. Atypical haemolytic uraemic syndrome with underlying glomerulopathies. A case series and a review of the literature. *Nephrol Dial Transplant,* 28**,** 2246-59.

MARIAN, A. J. 2012. Challenges in medical applications of whole exome/genome sequencing discoveries. *Trends Cardiovasc Med,* 22**,** 219-23.

MARINOZZI, M. C., VERGOZ, L., RYBKINE, T., NGO, S., BETTONI, S., PASHOV, A., CAYLA, M., TABARIN, F., JABLONSKI, M., HUE, C., SMITH, R. J., NORIS, M., HALBWACHS-MECARELLI, L., DONADELLI, R., FREMEAUX-BACCHI, V. & ROUMENINA, L. T. 2014. Complement factor B mutations in atypical hemolytic uremic syndrome-disease-relevant or benign? *J Am Soc Nephrol,* 25**,** 2053-65.

MARTINELLI, D., DEODATO, F. & DIONISI-VICI, C. 2011. Cobalamin C defect: natural history, pathophysiology, and treatment. *J Inherit Metab Dis,* 34**,** 127-35.

MATHIESON, P. W. 2012. The podocyte cytoskeleton in health and in disease. *Clinical Kidney Journal,* 5**,** 498-501.

MATSUI, H., HOZUMI, Y., TANAKA, T., OKADA, M., NAKANO, T., SUZUKI, Y., ISEKI, K., KAKEHATA, S., TOPHAM, M. K. & GOTO, K. 2014. Role of the N-terminal hydrophobic residues of DGKepsilon in targeting the endoplasmic reticulum. *Biochim Biophys Acta,* 1842**,** 1440-50.

MATSUMOTO, M., KOKAME, K., SOEJIMA, K., MIURA, M., HAYASHI, S., FUJII, Y., IWAI, A., ITO, E., TSUJI, Y., TAKEDA-SHITAKA, M., IWADATE, M., UMEYAMA, H., YAGI, H., ISHIZASHI, H., BANNO, F., NAKAGAKI, T., MIYATA, T. & FUJIMURA, Y. 2004. Molecular characterization of ADAMTS13 gene mutations in Japanese patients with Upshaw-Schulman syndrome. *Blood,* 103**,** 1305-10.

MCCARTHY, M. I., ABECASIS, G. R., CARDON, L. R., GOLDSTEIN, D. B., LITTLE, J., IOANNIDIS, J. P. A. & HIRSCHHORN, J. N. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet,* 9**,** 356-369.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S., DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research,* 20**,** 1297-1303.

MCMAHON, G. 2011. *Post-transplant HUS* [Online]. Renal Fellow Network. Available: http://renalfellow.blogspot.co.uk/2011/05/post-transplant-hus.html [Accessed 10.10.14.

MCNEARNEY, T., BALLARD, L., SEYA, T. & ATKINSON, J. P. 1989. Membrane cofactor protein of complement is present on human fibroblast, epithelial, and endothelial cells. *J Clin Invest,* 84**,** 538-45.

MCRAE, J. L., DUTHY, T. G., GRIGGS, K. M., ORMSBY, R. J., COWAN, P. J., CROMER, B. A., MCKINSTRY, W. J., PARKER, M. W., MURPHY, B. F. & GORDON, D. L. 2005. Human Factor H-Related Protein 5 Has Cofactor Activity, Inhibits C3 Convertase Activity, Binds Heparin and C-Reactive Protein, and Associates with Lipoprotein. *The Journal of Immunology,* 174**,** 6250-6256.

MCVEY, M. & LEE, S. E. 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in Genetics,* 24**,** 529-538.

MELE, C., LEMAIRE, M., IATROPOULOS, P., PIRAS, R., BRESIN, E., BETTONI, S., BICK, D., HELBLING, D., VEITH, R., VALOTI, E., DONADELLI, R., MURER, L., NEUNHAUSERER, M., BRENO, M., FREMEAUX-BACCHI, V., LIFTON, R., REMUZZI, G. & NORIS, M. 2015. Characterization of a New DGKE Intronic Mutation in Genetically Unsolved Cases of Familial Atypical Hemolytic Uremic Syndrome. *Clin J Am Soc Nephrol*.

MERI, S. 2013. Complement activation in diseases presenting with thrombotic microangiopathy. *European Journal of Internal Medicine,* 24**,** 496-502.

METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet,* 11**,** 31-46.

MEYNERT, A. M., ANSARI, M., FITZPATRICK, D. R. & TAYLOR, M. S. 2014. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics,* 15**,** 247.

MILDER, F. J., GOMES, L., SCHOUTEN, A., JANSSEN, B. J., HUIZINGA, E. G., ROMIJN, R. A., HEMRIKA, W., ROOS, A., DAHA, M. R. & GROS, P. 2007. Factor B structure provides insights into activation of the central protease of the complement system. *Nat Struct Mol Biol,* 14**,** 224-8.

MILLER, J. N. & PEARCE, D. A. 2014. Nonsense-mediated decay in genetic disease: Friend or foe? *Mutation Research/Reviews in Mutation Research,* 762**,** 52-64.

MOORE, I., STRAIN, L., PAPPWORTH, I., KAVANAGH, D., BARLOW, P. N., HERBERT, A. P., SCHMIDT, C. Q., STANIFORTH, S. J., HOLMES, L. V., WARD, R., MORGAN, L., GOODSHIP, T. H. & MARCHBANK, K. J. 2010. Association of factor H autoantibodies with deletions of CFHR1, CFHR3, CFHR4, and with mutations in CFH, CFI, CD46, and C3 in patients with atypical hemolytic uremic syndrome. *Blood,* 115**,** 379-87.

MOREL, C. F., LERNER-ELLIS, J. P. & ROSENBLATT, D. S. 2006. Combined methylmalonic aciduria and homocystinuria (cblC): Phenotype–genotype correlations and ethnic-specific observations. *Molecular Genetics and Metabolism,* 88**,** 315-321.

MORGAN, H. P., MERTENS, H. D., GUARIENTO, M., SCHMIDT, C. Q., SOARES, D. C., SVERGUN, D. I., HERBERT, A. P., BARLOW, P. N. & HANNAN, J. P. 2012. Structural analysis of the C-terminal region (modules 18-20) of complement regulator factor H (FH). *PLoS One,* 7**,** e32187.

MORGAN, H. P., SCHMIDT, C. Q., GUARIENTO, M., BLAUM, B. S., GILLESPIE, D., HERBERT, A. P., KAVANAGH, D., MERTENS, H. D., SVERGUN, D. I., JOHANSSON, C. M., UHRIN, D., BARLOW, P. N. & HANNAN, J. P. 2011. Structural basis for engagement by complement factor H of C3b on a self surface. *Nat Struct Mol Biol,* 18**,** 463-70.

MORLEY, B. J. & WALPORT, M. J. 2000. *The Complement FactsBook,* London, United Kingdom, Academic Press.

MRC-HOLLAND. 2013. MLPA General Protocol. Available: https://http://www.mlpa.com/WebForms/WebFormMain.aspx?Tag=_wl2zCji-rCGANQgZPuTixtCplCA1mmwJoFo_xHPnTgc. [Accessed 29.10.13].

MURER, L., ZACCHELLO, G., BIANCHI, D., DALL'AMICO, R., MONTINI, G., ANDREETTA, B., PERINI, M., DOSSI, E. C., ZANON, G. & ZACCHELLO, F. 2000. Thrombotic Microangiopathy Associated with Parvovirus B 19 Infection after Renal Transplantation. *Journal of the American Society of Nephrology,* 11**,** 1132-1137.

NAKANO, T., HOZUMI, Y., GOTO, K. & WAKABAYASHI, I. 2009. Localization of diacylglycerol kinase epsilon on stress fibers in vascular smooth muscle cells. *Cell Tissue Res,* 337**,** 167-75.

NANGAKU, M., ALPERS, C. E., PIPPIN, J., SHANKLAND, S. J., KUROKAWA, K., ADLER, S., MORGAN, B. P., JOHNSON, R. J. & COUSER, W. G. 1998. CD59 protects glomerular endothelial cells from immune-mediated thrombotic microangiopathy in rats. *J Am Soc Nephrol,* 9**,** 590-7.

NAUTA, A. J., DAHA, M. R., TIJSMA, O., VAN DE WATER, B., TEDESCO, F. & ROOS, A. 2002. The membrane attack complex of complement induces caspase activation and apoptosis. *Eur J Immunol,* 32**,** 783-92.

NCBI RESOURCE COORDINATORS 2014. Database resources of the National Center for Biotechnology Information. *Nucleic acids research*.

NG, S. B., BIGHAM, A. W., BUCKINGHAM, K. J., HANNIBAL, M. C., MCMILLIN, M. J., GILDERSLEEVE, H. I., BECK, A. E., TABOR, H. K., COOPER, G. M., MEFFORD, H. C., LEE, C., TURNER, E. H., SMITH, J. D., RIEDER, M. J., YOSHIURA, K.-I., MATSUMOTO, N., OHTA, T., NIIKAWA, N., NICKERSON, D. A., BAMSHAD, M. J. & SHENDURE, J. 2010a. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet,* 42**,** 790-793.

NG, S. B., BUCKINGHAM, K. J., LEE, C., BIGHAM, A. W., TABOR, H. K., DENT, K. M., HUFF, C. D., SHANNON, P. T., JABS, E. W., NICKERSON, D. A.,

SHENDURE, J. & BAMSHAD, M. J. 2010b. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet,* 42**,** 30-35.

NG, S. B., TURNER, E. H., ROBERTSON, P. D., FLYGARE, S. D., BIGHAM, A. W., LEE, C., SHAFFER, T., WONG, M., BHATTACHARJEE, A., EICHLER, E. E., BAMSHAD, M., NICKERSON, D. A. & SHENDURE, J. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature,* 461**,** 272-6.

NICE. 2015. *Eculizumab for treating atypical haemolytic uraemic syndrome* [Online]. Available: https://http://www.nice.org.uk/guidance/hst1 [Accessed 23.04.15.

NILSSON, S. C., KALCHISHKOVA, N., TROUW, L. A., FREMEAUX-BACCHI, V., VILLOUTREIX, B. O. & BLOM, A. M. 2010. Mutations in complement factor I as found in atypical hemolytic uremic syndrome lead to either altered secretion or altered function of factor I. *Eur J Immunol,* 40**,** 172-85.

NISHIGUCHI, K. M., YASUMA, T. R., TOMIDA, D., NAKAMURA, M., ISHIKAWA, K., KIKUCHI, M., OHMI, Y., NIWA, T., HAMAJIMA, N., FURUKAWA, K. & TERASAKI, H. 2012. C9-R95X Polymorphism in Patients with Neovascular Age-Related Macular Degeneration. *Investigative Ophthalmology & Visual Science,* 53**,** 508-512.

NISHIMURA, J., YAMAMOTO, M., HAYASHI, S., OHYASHIKI, K., ANDO, K., BRODSKY, A. L., NOJI, H., KITAMURA, K., ETO, T., TAKAHASHI, T., MASUKO, M., MATSUMOTO, T., WANO, Y., SHICHISHIMA, T., SHIBAYAMA, H., HASE, M., LI, L., JOHNSON, K., LAZAROWSKI, A., TAMBURINI, P., INAZAWA, J., KINOSHITA, T. & KANAKURA, Y. 2014. Genetic variants in C5 and poor response to eculizumab. *N Engl J Med,* 370**,** 632-9.

NORIS, M., CAPRIOLI, J., BRESIN, E., MOSSALI, C., PIANETTI, G., GAMBA, S., DAINA, E., FENILI, C., CASTELLETTI, F., SOROSINA, A., PIRAS, R., DONADELLI, R., MARANTA, R., VAN DER MEER, I., CONWAY, E. M., ZIPFEL, P. F., GOODSHIP, T. H. & REMUZZI, G. 2010. Relative Role of Genetic Complement Abnormalities in Sporadic and Familial aHUS and Their Impact on Clinical Phenotype. *Clinical Journal of the American Society of Nephrology,* 5**,** 1844-1859.

NORIS, M., MELE, C. & REMUZZI, G. 2015. Podocyte dysfunction in atypical haemolytic uraemic syndrome. *Nat Rev Nephrol*.

NORIS, M. & REMUZZI, G. 2009. Atypical hemolytic-uremic syndrome. *N Engl J Med,* 361**,** 1676-87.

NORIS, M. & REMUZZI, G. 2010. Thrombotic microangiopathy after kidney transplantation. *Am J Transplant,* 10**,** 1517-23.

NUNN, D. L. & WATSON, S. P. 1987. A diacylglycerol kinase inhibitor, R59022, potentiates secretion by and aggregation of thrombin-stimulated human platelets. *Biochem J,* 243**,** 809-13.

NURDEN, A. T. 2011. Platelets, inflammation and tissue regeneration. *Thrombosis and Haemostasis,* 105**,** S13-S33.

ORAN, A. E. & ISENMAN, D. E. 1999. Identification of residues within the 727-767 segment of human complement component C3 important for its interaction with factor H and with complement receptor 1 (CR1, CD35). *J Biol Chem,* 274**,** 5120-30.

OZALTIN, F., LI, B., RAUHAUSER, A., AN, S. W., SOYLEMEZOGLU, O., GONUL, II, TASKIRAN, E. Z., IBSIRLIOGLU, T., KORKMAZ, E., BILGINER, Y., DUZOVA, A., OZEN, S., TOPALOGLU, R., BESBAS, N., ASHRAF, S., DU, Y., LIANG, C., CHEN, P., LU, D., VADNAGARA, K., ARBUCKLE, S., LEWIS, D., WAKELAND, B., QUIGG, R. J., RANSOM, R. F., WAKELAND, E. K., TOPHAM, M. K., BAZAN, N. G., MOHAN, C., HILDEBRANDT, F., BAKKALOGLU, A., HUANG, C. L. & ATTANASIO, M. 2012. DGKE Variants Cause a Glomerular Microangiopathy That Mimics Membranoproliferative GN. *Journal of the American Society of Nephrology*.

PARK, H. J., KIM, H. J., HONG, Y. B., NAM, S. H., CHUNG, K. W. & CHOI, B. O. 2014. A novel INF2 mutation in a Korean family with autosomal dominant intermediate Charcot-Marie-Tooth disease and focal segmental glomerulosclerosis. *J Peripher Nerv Syst*.

PEREZ-CABALLERO, D., GONZALEZ-RUBIO, C., GALLARDO, M. E., VERA, M., LOPEZ-TRASCASA, M., RODRIGUEZ DE CORDOBA, S. & SANCHEZ-CORRAL, P. 2001. Clustering of missense mutations in the C-terminal region of factor H in atypical hemolytic uremic syndrome. *American Journal of Human Genetics,* 68**,** 478-84.

PERSSON, B. D., SCHMITZ, N. B., SANTIAGO, C., ZOCHER, G., LARVIE, M., SCHEU, U., CASASNOVAS, J. M. & STEHLE, T. 2010. Structure of the extracellular portion of CD46 provides insights into its interactions with complement proteins and pathogens. *PLoS Pathog,* 6**,** e1001122.

PEYVANDI, F. *TTP mutation database* [Online]. Luigi Villa Foundation. Available: http://www.ttpdatabase.org/en/table1.htm [Accessed 24.08.2014.

PEYVANDI, F., FERRARI, S., LAVORETANO, S., CANCIANI, M. T. & MANNUCCI, P. M. 2004. von Willebrand factor cleaving protease (ADAMTS-13) and ADAMTS-13 neutralizing autoantibodies in 100 patients with thrombotic thrombocytopenic purpura. *Br J Haematol,* 127**,** 433-9.

PICKERING, M. C., DE JORGE, E. G., MARTINEZ-BARRICARTE, R., RECALDE, S., GARCIA-LAYANA, A., ROSE, K. L., MOSS, J., WALPORT, M. J., COOK, H. T., DE CORDOBA, S. R. & BOTTO, M. 2007. Spontaneous hemolytic uremic syndrome triggered by complement factor H lacking surface recognition domains. *Journal of Experimental Medicine,* 204**,** 1249-56.

PIMANDA, J. E., MAEKAWA, A., WIND, T., PAXTON, J., CHESTERMAN, C. N. & HOGG, P. J. 2004. Congenital thrombotic thrombocytopenic purpura in association with a mutation in the second CUB domain of ADAMTS13. *Blood,* 103**,** 627-9.

PINARD, R., DE WINTER, A., SARKIS, G. J., GERSTEIN, M. B., TARTARO, K. R., PLANT, R. N., EGHOLM, M., ROTHBERG, J. M. & LEAMON, J. H. 2006. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics,* 7**,** 216.

PLAGNOL, V., CURTIS, J., EPSTEIN, M., MOK, K. Y., STEBBINGS, E., GRIGORIADOU, S., WOOD, N. W., HAMBLETON, S., BURNS, S. O., THRASHER,

A. J., KUMARARATNE, D., DOFFINGER, R. & NEJENTSEV, S. 2012. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics,* 28**,** 2747-54.

PLAIMAUER, B., FUHRMANN, J., MOHR, G., WERNHART, W., BRUNO, K., FERRARI, S., KONETSCHNY, C., ANTOINE, G., RIEGER, M. & SCHEIFLINGER, F. 2006. Modulation of ADAMTS13 secretion and specific activity by a combination of common amino acid polymorphisms and a missense mutation. *Blood,* 107**,** 118-25.

PODACK, E. R., TSCHOOP, J. & MULLER-EBERHARD, H. J. 1982. Molecular organization of C9 within the membrane attack complex of complement. Induction of circular C9 polymerization by the C5b-8 assembly. *The Journal of Experimental Medicine,* 156**,** 268-282.

POLLAK, M. R., QUAGGIN, S. E., HOENIG, M. P. & DWORKIN, L. D. 2014. The Glomerulus: The Sphere of Influence. *Clinical Journal of the American Society of Nephrology,* 9**,** 1461-1469.

POLLARD, K. S., HUBISZ, M. J., ROSENBLOOM, K. R. & SIEPEL, A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research,* 20**,** 110-21.

PONTICELLI, C. & BANFI, G. 2006. Thrombotic microangiopathy after kidney transplantation. *Transpl Int,* 19**,** 789-94.

PUERTOLLANO, R. & ALONSO, M. A. 1999. MAL, an integral element of the apical sorting machinery, is an itinerant protein that cycles between the trans-Golgi network and the plasma membrane. *Mol Biol Cell,* 10**,** 3435-47.

QIU, W., ZHANG, Y., LIU, X., ZHOU, J., LI, Y., ZHOU, Y., SHAN, K., XIA, M., CHE, N., FENG, X., ZHAO, D. & WANG, Y. 2012. Sublytic C5b-9 complexes induce proliferative changes of glomerular mesangial cells in rat Thy-1 nephritis through TRAF6-mediated PI3K-dependent Akt1 activation. *The Journal of Pathology,* 226**,** 619-632.

RAM, S., SHARMA, A. K., SIMPSON, S. D., GULATI, S., MCQUILLEN, D. P., PANGBURN, M. K. & RICE, P. A. 1998. A Novel Sialic Acid Binding Site on Factor H Mediates Serum Resistance of Sialylated Neisseria gonorrhoeae. *The Journal of Experimental Medicine,* 187**,** 743-752.

RAMABHADRAN, V., HATCH, A. L. & HIGGS, H. N. 2013. Actin monomers activate inverted formin 2 by competing with its autoinhibitory interaction. *J Biol Chem,* 288**,** 26847-55.

RAMABHADRAN, V., KOROBOVA, F., RAHME, G. J. & HIGGS, H. N. 2011. Splice variant-specific cellular function of the formin INF2 in maintenance of Golgi architecture. *Mol Biol Cell,* 22**,** 4822-33.

REID, K. B. M. & DAY, A. J. 1989. Structure-function relationships of the complement components. *Immunology Today,* 10**,** 177-180.

REN, S., SHATADAL, S. & SHEN, G. X. 2000. Protein kinase C-beta mediates lipoprotein-induced generation of PAI-1 from vascular endothelial cells. *Am J Physiol Endocrinol Metab,* 278**,** E656-62.

REVA, B., ANTIPIN, Y. & SANDER, C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research*.

RICHARDS, A., BUDDLES, M. R., DONNE, R. L., KAPLAN, B. S., KIRK, E., VENNING, M. C., TIELEMANS, C. L., GOODSHIP, J. A. & GOODSHIP, T. H. 2001. Factor H mutations in hemolytic uremic syndrome cluster in exons 18-20, a domain important for host cell recognition. *American Journal of Human Genetics,* 68**,** 485-90.

RICHARDS, A., KEMP, E. J., LISZEWSKI, M. K., GOODSHIP, J. A., LAMPE, A. K., DECORTE, R., MUSLUMANOGLU, M. H., KAVUKCU, S., FILLER, G., PIRSON, Y., WEN, L. S., ATKINSON, J. P. & GOODSHIP, T. H. 2003. Mutations in human complement regulator, membrane cofactor protein (CD46), predispose to development of familial hemolytic uremic syndrome. *Proc Natl Acad Sci U S A,* 100**,** 12966-71.

RICKLIN, D., HAJISHENGALLIS, G., YANG, K. & LAMBRIS, J. D. 2010. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol,* 11**,** 785-97.

RIORDAN, J. R., ROMMENS, J. M., KEREM, B., ALON, N., ROZMAHEL, R., GRZELCZAK, Z., ZIELENSKI, J., LOK, S., PLAVSIC, N., CHOU, J. L. & ET AL. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science,* 245**,** 1066-73.

RIVIERE, J. B., MIRZAA, G. M., O'ROAK, B. J., BEDDAOUI, M., ALCANTARA, D., CONWAY, R. L., ST-ONGE, J., SCHWARTZENTRUBER, J. A., GRIPP, K. W., NIKKEL, S. M., WORTHYLAKE, T., SULLIVAN, C. T., WARD, T. R., BUTLER, H. E., KRAMER, N. A., ALBRECHT, B., ARMOUR, C. M., ARMSTRONG, L., CALUSERIU, O., CYTRYNBAUM, C., DROLET, B. A., INNES, A. M., LAUZON, J. L., LIN, A. E., MANCINI, G. M., MESCHINO, W. S., REGGIN, J. D., SAGGAR, A. K., LERMAN-SAGIE, T., UYANIK, G., WEKSBERG, R., ZIRN, B., BEAULIEU, C. L., MAJEWSKI, J., BULMAN, D. E., O'DRISCOLL, M., SHENDURE, J., GRAHAM, J. M., JR., BOYCOTT, K. M. & DOBYNS, W. B. 2012. De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat Genet,* 44**,** 934-40.

RIZZO, J. M. & BUCK, M. J. 2012. Key Principles and Clinical Applications of "Next-Generation" DNA Sequencing. *Cancer Prevention Research,* 5**,** 887-900.

ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nat Biotech,* 29**,** 24-26.

RODRIGUEZ DE TURCO, E. B., TANG, W., TOPHAM, M. K., SAKANE, F., MARCHESELLI, V. L., CHEN, C., TAKETOMI, A., PRESCOTT, S. M. & BAZAN, N. G. 2001. Diacylglycerol kinase epsilon regulates seizure susceptibility and long-term potentiation through arachidonoyl- inositol lipid signaling. *Proc Natl Acad Sci U S A,* 98**,** 4740-5.

RODRIGUEZ, E., RALLAPALLI, P. M., OSBORNE, A. & PERKINS, S. J. 2014. New functional and structural insights from updated mutational databases for complement factor H, factor I, membrane cofactor protein and C3. *Biosci Rep*.

RODRIGUEZ, P. Q., LOHKAMP, B., CELSI, G., MACHE, C. J., AUER-GRUMBACH, M., WERNERSON, A., HAMAJIMA, N., TRYGGVASON, K. & PATRAKKA, J. 2013. Novel INF2 mutation p. L77P in a family with glomerulopathy and Charcot-Marie-Tooth neuropathy. *Pediatric Nephrology,* 28**,** 339-43.

ROODHOOFT, A. M., MCLEAN, R. H., ELST, E. & VAN ACKER, K. J. 1990. Recurrent haemolytic uraemic syndrome and acquired hypomorphic variant of the third component of complement. *Pediatr Nephrol,* 4**,** 597-9.

ROSE, R., WEYAND, M., LAMMERS, M., ISHIZAKI, T., AHMADIAN, M. R. & WITTINGHOFER, A. 2005. Structural and mechanistic insights into the interaction between Rho and mammalian Dia. *Nature,* 435**,** 513-8.

ROTHER, R. P., ROLLINS, S. A., MOJCIK, C. F., BRODSKY, R. A. & BELL, L. 2007. Discovery and development of the complement inhibitor eculizumab for the treatment of paroxysmal nocturnal hemoglobinuria. *Nat Biotechnol,* 25**,** 1256-64.

ROUMENINA, L. T., FRIMAT, M., MILLER, E. C., PROVOT, F., DRAGON-DUREY, M. A., BORDEREAU, P., BIGOT, S., HUE, C., SATCHELL, S. C., MATHIESON, P. W., MOUSSON, C., NOEL, C., SAUTES-FRIDMAN, C., HALBWACHS-MECARELLI, L., ATKINSON, J. P., LIONET, A. & FREMEAUX-BACCHI, V. 2012. A prevalent C3 mutation in aHUS patients causes a direct C3 convertase gain of function. *Blood,* 119**,** 4182-91.

ROVERSI, P., JOHNSON, S., CAESAR, J. J. E., MCLEAN, F., LEATH, K. J., TSIFTSOGLOU, S. A., MORGAN, B. P., HARRIS, C. L., SIM, R. B. & LEA, S. M. 2011. Structural basis for complement factor I control and its disease-associated sequence polymorphisms. *Proceedings of the National Academy of Sciences*.

RUBNITZ, J. & SUBRAMANI, S. 1984. The minimum amount of homology required for homologous recombination in mammalian cells. *Molecular and Cellular Biology,* 4**,** 2253-2258.

RUGGENENTI, P. 2002. Post-transplant hemolytic-uremic syndrome. *Kidney Int,* 62**,** 1093-104.

RYKALINA, V. N., SHADRIN, A. A., AMSTISLAVSKIY, V. S., ROGAEV, E. I., LEHRACH, H. & BORODINA, T. A. 2014. Exome sequencing from nanogram amounts of starting DNA: comparing three approaches. *PLoS One,* 9**,** e101154.

SADLER, J. E. 1991. von Willebrand factor. *J Biol Chem,* 266**,** 22777-80.

SAKANE, F., KAI, M., WADA, I., IMAI, S. & KANOH, H. 1996. The C-terminal part of diacylglycerol kinase alpha lacking zinc fingers serves as a catalytic domain. *Biochemical Journal,* 318**,** 583-590.

SANCHEZ CHINCHILLA, D., PINTO, S., HOPPE, B., ADRAGNA, M., LOPEZ, L., JUSTA ROLDAN, M. L., PENA, A., LOPEZ TRASCASA, M., SANCHEZ-CORRAL, P. & RODRIGUEZ DE CORDOBA, S. 2014. Complement Mutations in Diacylglycerol Kinase-epsilon-Associated Atypical Hemolytic Uremic Syndrome. *Clin J Am Soc Nephrol*.

SANCHEZ-ARES, M., GARCIA-VIDAL, M., ANTUCHO, E. E., JULIO, P., EDUARDO, V. M., LENS, X. M. & GARCIA-GONZALEZ, M. A. 2013. A novel

mutation, outside of the candidate region for diagnosis, in the inverted formin 2 gene can cause focal segmental glomerulosclerosis. *Kidney International,* 83**,** 153-9.

SANCHEZ-GALLEGO, J. I., GROENEVELD, T. W. L., KRENTZ, S., NILSSON, S. C., VILLOUTREIX, B. O. & BLOM, A. M. 2012. Analysis of Binding Sites on Complement Factor I Using Artificial N-Linked Glycosylation. *Journal of Biological Chemistry,* 287**,** 13572-13583.

SARMA, J. V. & WARD, P. A. 2011. The complement system. *Cell Tissue Res,* 343**,** 227-35.

SCHMIDT, C. Q., HERBERT, A. P., HOCKING, H. G., UHRIN, D. & BARLOW, P. N. 2008a. Translational mini-review series on complement factor H: structural and functional correlations for factor H. *Clin Exp Immunol,* 151**,** 14-24.

SCHMIDT, C. Q., HERBERT, A. P., KAVANAGH, D., GANDY, C., FENTON, C. J., BLAUM, B. S., LYON, M., UHRIN, D. & BARLOW, P. N. 2008b. A new map of glycosaminoglycan and C3b binding sites on factor H. *J Immunol,* 181**,** 2610-9.

SCHNEPPENHEIM, R., BUDDE, U., OYEN, F., ANGERHAUS, D., AUMANN, V., DREWKE, E., HASSENPFLUG, W., HABERLE, J., KENTOUCHE, K., KOHNE, E., KURNIK, K., MUELLER-WIEFEL, D., OBSER, T., SANTER, R. & SYKORA, K. W. 2003. von Willebrand factor cleaving protease and ADAMTS13 mutations in childhood TTP. *Blood,* 101**,** 1845-50.

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.

SCHRAMM, E. C., CLARK, S. J., TRIEBWASSER, M. P., RAYCHAUDHURI, S., SEDDON, J. M. & ATKINSON, J. P. 2014. Genetic variants in the complement system predisposing to age-related macular degeneration: A review. *Molecular Immunology,* 61**,** 118-125.

SCHRAMM, E. C., ROUMENINA, L. T., RYBKINE, T., CHAUVET, S., VIEIRA-MARTINS, P., HUE, C., MAGA, T., VALOTI, E., WILSON, V., JOKIRANTA, S., SMITH, R. J., NORIS, M., GOODSHIP, T., ATKINSON, J. P. & FREMEAUX-BACCHI, V. 2015. Functional mapping of the interactions between complement C3 and regulatory proteins using atypical hemolytic uremic syndrome-associated mutations. *Blood*.

SCHRÖDINGER, L. *The PyMOL Molecular Graphics System* [Online]. Available: http://www.pymol.org/ [Accessed 29.01.15.

SCHWARZ, J. M., COOPER, D. N., SCHUELKE, M. & SEELOW, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Meth,* 11**,** 361-362.

SCHWARZ, J. M., RODELSPERGER, C., SCHUELKE, M. & SEELOW, D. 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Meth,* 7**,** 575-576.

SCULLY, M. & GOODSHIP, T. 2014. How I treat thrombotic thrombocytopenic purpura and atypical haemolytic uraemic syndrome. *Br J Haematol,* 164**,** 759-66.

SEDDON, J. M., YU, Y., MILLER, E. C., REYNOLDS, R., TAN, P. L., GOWRISANKAR, S., GOLDSTEIN, J. I., TRIEBWASSER, M., ANDERSON, H. E., ZERBIB, J., KAVANAGH, D., SOUIED, E., KATSANIS, N., DALY, M. J., ATKINSON, J. P. & RAYCHAUDHURI, S. 2013. Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nature Genetics,* 45**,** 1366-70.

SELLIER-LECLERC, A. L., FREMEAUX-BACCHI, V., DRAGON-DUREY, M. A., MACHER, M. A., NIAUDET, P., GUEST, G., BOUDAILLIEZ, B., BOUISSOU, F., DESCHENES, G., GIE, S., TSIMARATOS, M., FISCHBACH, M., MORIN, D., NIVET, H., ALBERTI, C. & LOIRAT, C. 2007. Differential impact of complement mutations on clinical characteristics in atypical hemolytic uremic syndrome. *Journal of the American Society of Nephrology,* 18**,** 2392-400.

SENIOR, M. & WALLACE, M. 2014. Fluorescence Imaging of MACPF/CDC Proteins: New Techniques and Their Application. *In:* ANDERLUH, G. & GILBERT, R. (eds.) *MACPF/CDC Proteins - Agents of Defence, Attack and Invasion.* Springer Netherlands.

SEYA, T., TURNER, J. R. & ATKINSON, J. P. 1986. Purification and characterization of a membrane protein (gp45-70) that is a cofactor for cleavage of C3b and C4b. *J Exp Med,* 163**,** 837-55.

SHAPIRO, M. B. & SENAPATHY, P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res,* 15**,** 7155-74.

SHARMA, A. K. & PANGBURN, M. K. 1996. Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Proc Natl Acad Sci U S A,* 93**,** 10996-1001.

SHARMA, A. P., GREENBERG, C. R., PRASAD, A. N. & PRASAD, C. 2007. Hemolytic uremic syndrome (HUS) secondary to cobalamin C (cblC) disorder. *Pediatr Nephrol,* 22**,** 2097-103.

SHARMA, S., JAVADEKAR, S. M., PANDEY, M., SRIVASTAVA, M., KUMARI, R. & RAGHAVAN, S. C. 2015. Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell Death Dis,* 6**,** e1697.

SHAW, D. J., BROOK, J. D., MEREDITH, A. L., HARLEY, H. G., SARFARAZI, M. & HARPER, P. S. 1986. Gene mapping and chromosome 19. *J Med Genet,* 23**,** 2-10.

SHEERIN, N., KAVANAGH, D., GOODSHIP, T. H. J. & JOHNSON, S. 2015. A national specialised service in England for atypical haemolytic uraemic syndrome – the first year's experience. *QJM*.

SHENKMAN, B. & EINAV, Y. 2014. Thrombotic thrombocytopenic purpura and other thrombotic microangiopathic hemolytic anemias: diagnosis and classification. *Autoimmun Rev,* 13**,** 584-6.

SHERRY, S. T., WARD, M. H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E. M. & SIROTKIN, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res,* 29**,** 308-11.

SHIH, N. Y., LI, J., KARPITSKII, V., NGUYEN, A., DUSTIN, M. L., KANAGAWA, O., MINER, J. H. & SHAW, A. S. 1999. Congenital nephrotic syndrome in mice lacking CD2-associated protein. *Science,* 286**,** 312-5.

SHIHAB, H. A., GOUGH, J., COOPER, D. N., DAY, I. N. & GAUNT, T. R. 2013a. Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics,* 29**,** 1504-10.

SHIHAB, H. A., GOUGH, J., COOPER, D. N., STENSON, P. D., BARKER, G. L., EDWARDS, K. J., DAY, I. N. & GAUNT, T. R. 2013b. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat,* 34**,** 57-65.

SHIHAB, H. A., GOUGH, J., MORT, M., COOPER, D. N., DAY, I. N. & GAUNT, T. R. 2014. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Hum Genomics,* 8**,** 11.

SHULGA, Y. V., TOPHAM, M. K. & EPAND, R. M. 2011a. Regulation and functions of diacylglycerol kinases. *Chem Rev,* 111**,** 6186-208.

SHULGA, Y. V., TOPHAM, M. K. & EPAND, R. M. 2011b. Study of arachidonoyl specificity in two enzymes of the PI cycle. *J Mol Biol,* 409**,** 101-12.

SIEGLER, R. & OAKES, R. 2005. Hemolytic uremic syndrome; pathogenesis, treatment, and outcome. *Current Opinion in Pediatrics,* 17**,** 200-4.

SINGH, A. J., MEYER, R. D., BAND, H. & RAHIMI, N. 2005. The carboxyl terminus of VEGFR-2 is required for PKC-mediated down-regulation. *Mol Biol Cell,* 16**,** 2106-18.

SKERKA, C., CHEN, Q., FREMEAUX-BACCHI, V. & ROUMENINA, L. T. 2013. Complement factor H related proteins (CFHRs). *Molecular Immunology*.

SKERKA, C., KUHN, S., GUNTHER, K., LINGELBACH, K. & ZIPFEL, P. F. 1993. A novel short consensus repeat-containing molecule is related to human complement factor H. *J Biol Chem,* 268**,** 2904-8.

SMIGIELSKI, E. M., SIROTKIN, K., WARD, M. & SHERRY, S. T. 2000. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res,* 28**,** 352-5.

STAMENKOVIC, I. 2003. Extracellular matrix remodelling: the role of matrix metalloproteinases. *J Pathol,* 200**,** 448-64.

STOCKSCHLAEDER, M., SCHNEPPENHEIM, R. & BUDDE, U. 2014. Update on von Willebrand factor multimers: focus on high-molecular-weight multimers and their role in hemostasis. *Blood Coagul Fibrinolysis,* 25**,** 206-16.

STUPPIA, L., ANTONUCCI, I., PALKA, G. & GATTA, V. 2012. Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases. *Int J Mol Sci,* 13**,** 3245-76.

SULLIVAN, M., ERLIC, Z., HOFFMANN, M. M., ARBEITER, K., PATZER, L., BUDDE, K., HOPPE, B., ZEIER, M., LHOTTA, K., RYBICKI, L. A., BOCK, A., BERISHA, G. & NEUMANN, H. P. 2010. Epidemiological approach to identifying

genetic predispositions for atypical hemolytic uremic syndrome. *Ann Hum Genet,* 74**,** 17-26.

SUN, X., FUNK, C. D., DENG, C., SAHU, A., LAMBRIS, J. D. & SONG, W.-C. 1999. Role of decay-accelerating factor in regulating complement activation on the erythrocyte surface as revealed by gene targeting. *Proceedings of the National Academy of Sciences,* 96**,** 628-633.

SWIATECKA-URBAN, A. 2013. Membrane trafficking in podocyte health and disease. *Pediatr Nephrol,* 28**,** 1723-37.

SZOSTAK, J. W., ORR-WEAVER, T. L., ROTHSTEIN, R. J. & STAHL, F. W. 1983. The double-strand-break repair model for recombination. *Cell,* 33**,** 25-35.

TAKANO, T., CYBULSKY, A. V., YANG, X. & AOUDJIT, L. 2001. Complement C5b-9 induces cyclooxygenase-2 gene transcription in glomerular epithelial cells. *Am J Physiol Renal Physiol,* 281**,** F841-50.

TAKEDA, J., MIYATA, T., KAWAGOE, K., IIDA, Y., ENDO, Y., FUJITA, T., TAKAHASHI, M., KITANI, T. & KINOSHITA, T. 1993. Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria. *Cell,* 73**,** 703-11.

TANG, W., BUNTING, M., ZIMMERMAN, G. A., MCINTYRE, T. M. & PRESCOTT, S. M. 1996. Molecular cloning of a novel human diacylglycerol kinase highly selective for arachidonate-containing substrates. *J Biol Chem,* 271**,** 10237-41.

TAO, Z., ANTHONY, K., PENG, Y., CHOI, H., NOLASCO, L., RICE, L., MOAKE, J. L. & DONG, J. F. 2006. Novel ADAMTS-13 mutations in an adult with delayed onset thrombotic thrombocytopenic purpura. *J Thromb Haemost,* 4**,** 1931-5.

TARR, P. I., GORDON, C. A. & CHANDLER, W. L. 2005. Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. *The Lancet,* 365**,** 1073-1086.

TAWADROUS, H., MAGA, T., SHARMA, J., KUPFERMAN, J., SMITH, R. & SCHOENEMAN, M. 2010. A novel mutation in the Complement Factor B gene (CFB) and atypical hemolytic uremic syndrome. *Pediatric Nephrology,* 25**,** 947-951.

TAYLOR, C. M., MACHIN, S., WIGMORE, S. J., GOODSHIP, T. H. J., ON BEHALF OF A WORKING PARTY FROM THE RENAL ASSOCIATION, T. B. C. F. S. I. H. & THE BRITISH TRANSPLANTATION, S. 2010. Clinical Practice Guidelines for the management of atypical Haemolytic Uraemic Syndrome in the United Kingdom. *British Journal of Haematology,* 148**,** 37-47.

TEGLA, C., CUDRICI, C., PATEL, S., TRIPPE, R., III, RUS, V., NICULESCU, F. & RUS, H. 2011. Membrane attack by complement: the assembly and biology of terminal complement complexes. *Immunologic Research,* 51**,** 45-60.

TENNESSEN, J. A., BIGHAM, A. W., O'CONNOR, T. D., FU, W., KENNY, E. E., GRAVEL, S., MCGEE, S., DO, R., LIU, X., JUN, G., KANG, H. M., JORDAN, D., LEAL, S. M., GABRIEL, S., RIEDER, M. J., ABECASIS, G., ALTSHULER, D., NICKERSON, D. A., BOERWINKLE, E., SUNYAEV, S., BUSTAMANTE, C. D., BAMSHAD, M. J. & AKEY, J. M. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science,* 337**,** 64-9.

THERMO-SCIENTIFIC. Crosslinking Reagents Technical Handbook. Available: https://tools.lifetechnologies.com/content/sfs/brochures/1602163-Crosslinking-Reagents-Handbook.pdf [Accessed 24.03.15].

THERMO-SCIENTIFIC. 2013. *Instructions 1296.7: Pierce™ BCA Protein Assay Kit* [Online]. Available: http://www.piercenet.com/pisearch/?prodnum=23225 [Accessed 18.03.13.

THOMAS, M. L., JANATOVA, J., GRAY, W. R. & TACK, B. F. 1982. Third component of human complement: localization of the internal thiolester bond. *Proc Natl Acad Sci U S A,* 79**,** 1054-8.

THOMPSON, R. A. & WINTERBORN, M. H. 1981. Hypocomplementaemia due to a genetic deficiency of beta 1H globulin. *Clin Exp Immunol,* 46**,** 110-9.

THORVALDSDÓTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics,* 14**,** 178-192.

TOPHAM, M. K. & EPAND, R. M. 2009. Mammalian diacylglycerol kinases: Molecular interactions and biological functions of selected isoforms. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1790**,** 416-424.

TORREIRA, E., TORTAJADA, A., MONTES, T., RODRIGUEZ DE CORDOBA, S. & LLORCA, O. 2009. Coexistence of closed and open conformations of complement factor B in the alternative pathway C3bB(Mg2+) proconvertase. *J Immunol,* 183**,** 7347-51.

TORTAJADA, A., MONTES, T., MARTINEZ-BARRICARTE, R., MORGAN, B. P., HARRIS, C. L. & DE CORDOBA, S. R. 2009. The disease-protective complement factor H allotypic variant Ile62 shows increased binding affinity for C3b and enhanced cofactor activity. *Hum Mol Genet,* 18**,** 3452-61.

TOYOTA, K., OGINO, D., HAYASHI, M., TAKI, M., SAITO, K., ABE, A., HASHIMOTO, T., UMETSU, K., TSUKAGUCHI, H. & HAYASAKA, K. 2013. INF2 mutations in Charcot-Marie-Tooth disease complicated with focal segmental glomerulosclerosis. *J Peripher Nerv Syst,* 18**,** 97-8.

TROUW, L. A. & DAHA, M. R. 2011. Role of complement in innate immunity and host defense. *Immunol Lett,* 138**,** 35-7.

TSIANG, M., LENTZ, S. R. & SADLER, J. E. 1992. Functional domains of membrane-bound human thrombomodulin. EGF-like domains four to six and the serine/threonine-rich domain are required for cofactor activity. *J Biol Chem,* 267**,** 6164-70.

VALOTI, E., ALBERTI, M., TORTAJADA, A., GARCIA-FERNANDEZ, J., GASTOLDI, S., BESSO, L., BRESIN, E., REMUZZI, G., RODRIGUEZ DE CORDOBA, S. & NORIS, M. 2015. A novel atypical hemolytic uremic syndrome-associated hybrid CFHR1/CFH gene encoding a fusion protein that antagonizes factor H-dependent complement regulation. *J Am Soc Nephrol,* 26**,** 209-19.

VAN DER AUWERA, G. A., CARNEIRO, M. O., HARTL, C., POPLIN, R., DEL ANGEL, G., LEVY-MOONSHINE, A., JORDAN, T., SHAKIR, K., ROAZEN, D.,

THIBAULT, J., BANKS, E., GARIMELLA, K. V., ALTSHULER, D., GABRIEL, S. & DEPRISTO, M. A. 2002. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics.* John Wiley & Sons, Inc.

VENABLES, J. P., STRAIN, L., ROUTLEDGE, D., BOURN, D., POWELL, H. M., WARWICKER, P., DIAZ-TORRES, M. L., SAMPSON, A., MEAD, P., WEBB, M., PIRSON, Y., JACKSON, M. S., HUGHES, A., WOOD, K. M., GOODSHIP, J. A. & GOODSHIP, T. H. 2006. Atypical haemolytic uraemic syndrome associated with a hybrid complement gene. *PLoS Med,* 3**,** e431.

VEYRADIER, A., LAVERGNE, J. M., RIBBA, A. S., OBERT, B., LOIRAT, C., MEYER, D. & GIRMA, J. P. 2004. Ten candidate ADAMTS13 mutations in six French families with congenital thrombotic thrombocytopenic purpura (Upshaw-Schulman syndrome). *J Thromb Haemost,* 2**,** 424-9.

VIJAYAN, K. V. 2015. DGKE disruption ditches complement and drives p38 signaling. *Blood,* 125**,** 898-9.

VYSE, T. J., BATES, G. P., WALPORT, M. J. & MORLEY, B. J. 1994. The organization of the human complement factor I gene (IF): a member of the serine protease gene family. *Genomics,* 24**,** 90-8.

WALPORT, M. J. 2001a. Complement. First of two parts. *N Engl J Med,* 344**,** 1058-66.

WALPORT, M. J. 2001b. Complement. Second of two parts. *N Engl J Med,* 344**,** 1140-4.

WANG, K., LI, M. & HAKONARSON, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research,* 38**,** e164-e164.

WARWICKER, P., GOODSHIP, T. H., DONNE, R. L., PIRSON, Y., NICHOLLS, A., WARD, R. M., TURNPENNY, P. & GOODSHIP, J. A. 1998. Genetic studies into inherited and sporadic hemolytic uremic syndrome. *Kidney International,* 53**,** 836-44.

WATANABE, H., GOTO, S., YAMAZAKI, H., YAMAMOTO, T. & NARITA, I. Exome Sequencing Identified Novel Mutations of FKRP in Familial Atypical Hemolytic Uremic Syndrome Kidney Week 2014, 13.11.14 2014 Philadelphia, PA, USA. American Society of Nephrology, 173.

WEN, D. Z., DITTMAN, W. A., YE, R. D., DEAVEN, L. L., MAJERUS, P. W. & SADLER, J. E. 1987. Human thrombomodulin: complete cDNA sequence and chromosome localization of the gene. *Biochemistry,* 26**,** 4350-7.

WESTRA, D., VOLOKHINA, E., VAN DER HEIJDEN, E., VOS, A., HUIGEN, M., JANSEN, J., VAN KAAUWEN, E., VAN DER VELDEN, T., VAN DE KAR, N. & VAN DEN HEUVEL, L. 2010. Genetic disorders in complement (regulating) genes in patients with atypical haemolytic uraemic syndrome (aHUS). *Nephrol Dial Transplant,* 25**,** 2195-202.

WHALEY, K. & NORTH, J. 1997. Haemolytic assays for whole complement activity and individual components. *In:* A. W. DODDS & R. B. SIM (eds.) *Complement A Practical Approach.* New York, USA: Oxford University Press inc.

WIEDMER, T., ESMON, C. T. & SIMS, P. J. 1986. Complement proteins C5b-9 stimulate procoagulant activity through platelet prothrombinase. *Blood,* 68**,** 875-80.

WINN, M. P., CONLON, P. J., LYNN, K. L., FARRINGTON, M. K., CREAZZO, T., HAWKINS, A. F., DASKALAKIS, N., KWAN, S. Y., EBERSVILLER, S., BURCHETTE, J. L., PERICAK-VANCE, M. A., HOWELL, D. N., VANCE, J. M. & ROSENBERG, P. B. 2005. A Mutation in the TRPC6 Cation Channel Causes Familial Focal Segmental Glomerulosclerosis. *Science,* 308**,** 1801-1804.

WITZEL-SCHLOMP, K., SPATH, P. J., HOBART, M. J., FERNIE, B. A., RITTNER, C., KAUFMANN, T. & SCHNEIDER, P. M. 1997. The human complement C9 gene: identification of two mutations causing deficiency and revision of the gene structure. *J Immunol,* 158**,** 5043-9.

WONG, E. K., ANDERSON, H. E., HERBERT, A. P., CHALLIS, R. C., BROWN, P., REIS, G. S., TELLEZ, J. O., STRAIN, L., FLUCK, N., HUMPHREY, A., MACLEOD, A., RICHARDS, A., AHLERT, D., SANTIBANEZ-KOREF, M., BARLOW, P. N., MARCHBANK, K. J., HARRIS, C. L., GOODSHIP, T. H. & KAVANAGH, D. 2014. Characterization of a Factor H Mutation That Perturbs the Alternative Pathway of Complement in a Family with Membranoproliferative GN. *Journal of the American Society of Nephrology*.

WONG, E. K. & KAVANAGH, D. 2015. Anticomplement C5 therapy with eculizumab for the treatment of paroxysmal nocturnal hemoglobinuria and atypical hemolytic uremic syndrome. *Transl Res,* 165**,** 306-320.

WONG, E. K. S., GOODSHIP, T. H. J. & KAVANAGH, D. 2013. Complement therapy in atypical haemolytic uraemic syndrome (aHUS). *Molecular Immunology,* 56**,** 199-212.

XU, H., LUO, X., QIAN, J., PANG, X., SONG, J., QIAN, G., CHEN, J. & CHEN, S. 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One,* 7**,** e52249.

YADA, Y., OZEKI, T., KANOH, H. & NOZAWA, Y. 1990. Purification and characterization of cytosolic diacylglycerol kinases of human platelets. *J Biol Chem,* 265**,** 19237-43.

YONEMURA, Y., KAWAKITA, M., KOITO, A., KAWAGUCHI, T., NAKAKUMA, H., KAGIMOTO, T., SCHICHISHIMA, T., TERASAWA, T., AKAGAKI, Y. & INAI, S. 1990. Paroxysmal nocturnal haemoglobinuria with coexisting deficiency of the ninth component of complement: lack of massive haemolytic attack. *Br J Haematol,* 74**,** 108-13.

YU, G. H., HOLERS, V. M., SEYA, T., BALLARD, L. & ATKINSON, J. P. 1986. Identification of a third component of complement-binding glycoprotein of human platelets. *J Clin Invest,* 78**,** 494-501.

ZARIFIAN, A., MELEG-SMITH, S., O'DONOVAN, R., TESI, R. J. & BATUMAN, V. 1999. Cyclosporine-associated thrombotic microangiopathy in renal allografts. *Kidney Int,* 55**,** 2457-66.

ZHENG, X. L., WU, H. M., SHANG, D., FALLS, E., SKIPWITH, C. G., CATALAND, S. R., BENNETT, C. L. & KWAAN, H. C. 2010. Multiple domains of ADAMTS13 are targeted by autoantibodies against ADAMTS13 in patients with acquired idiopathic thrombotic thrombocytopenic purpura. *Haematologica,* 95**,** 1555-1562.

ZIPFEL, P. F., EDEY, M., HEINEN, S., JOZSI, M., RICHTER, H., MISSELWITZ, J., HOPPE, B., ROUTLEDGE, D., STRAIN, L., HUGHES, A. E., GOODSHIP, J. A., LICHT, C., GOODSHIP, T. H. & SKERKA, C. 2007. Deletion of complement factor H-related genes CFHR1 and CFHR3 is associated with atypical hemolytic uremic syndrome. *Plos Genetics,* 3**,** e41.

ZIPFEL, P. F. & SKERKA, C. 1999. FHL-1/reconectin: a human complement and immune regulator with cell-adhesive function. *Immunol Today,* 20**,** 135-40.

ZIPFEL, P. F. & SKERKA, C. 2001. Complement: The Alternative Pathway. *eLS.* John Wiley & Sons, Ltd.

ZOPPI, M., WEISS, M., NYDEGGER, U. E., HESS, T. & SPATH, P. J. 1990. Recurrent meningitis in a patient with congenital deficiency of the C9 component of complement. First case of C9 deficiency in Europe. *Arch Intern Med,* 150**,** 2395-9.