Genetic Investigation of Kidney Disease

A thesis submitted for the Degree of Doctor of Philosophy

2010

Daniel P. Gale
UCL Division of Medicine

Declaration

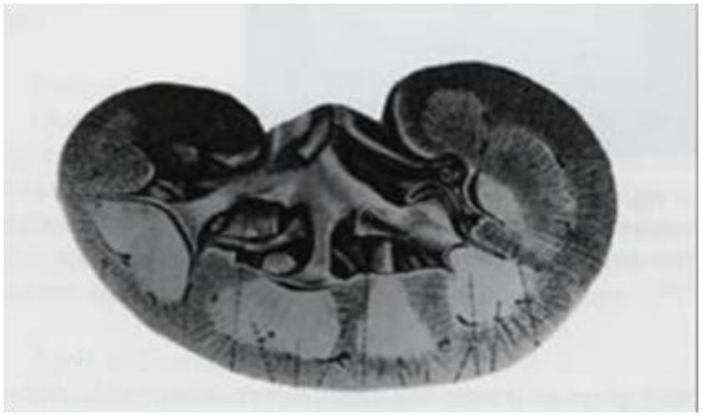
I, Daniel Philip Gale, confirm that the work presented in this thesis is my own. Where information has been derived from other sources I confirm that this has been indicated in the thesis.

Acknowledgements

I would like to thank the patients and families who participated in this study and from whom I have learned so much. Without their trust, hospitality and generosity none of this would have been possible. I am especially glad to have the opportunity to thank my supervisor, Professor Patrick Maxwell whose support, encouragement and wisdom have been so valuable in both the laboratory and the hospital. I would also like to thank the Medical Research Council for funding this work.

I also owe a debt of gratitude to my colleagues in the laboratories and hospitals in which I worked – especially Drs Sarah Harten, Tapan Bhattacharyya, Alkis Pierides, Margaret Ashcroft and Deepa Shukla; and to my collaborators, in particular Professors Terry Cook, Ted Tuddenham and Peter Robbins, and Drs Matthew Pickering, Nick Talbot and Elena Goicoechea de Jorge.

I would also like to thank Guy Neild, Tom Connor, Tony Segal and Oliver Staples for many stimulating discussions, and my family and friends for their unwavering support and interest.



Abstract

Kidney disease is an important contributor to the burden of ill-health worldwide. Genetic factors have an important role in determining which people are affected by kidney disease, and this study aimed to identify the genes responsible for disease in families in which unusual kidney diseases were transmitted in a pattern suggesting autosomal dominant inheritance. I performed genome-wide single nucleotide polymorphism-based linkage studies and identified two new human disease genes.

Hypoxia Inducible Factor-2α (HIF2α) is a widely expressed transcription factor which is rapidly broken down in the presence of oxygen. When oxygenation is reduced it activates the transcription of many genes, including *erythropoietin* which stimulates red blood cell production. I identified a heterozygous activating mutation of *HIF2α* which cosegregated with autosomal dominantly inherited erythrocytosis and pulmonary arterial hypertension in a British family, producing a phenotype similar to the effects of high altitude exposure. *In vitro* studies demonstrated that the mutant protein has increased transcriptional activity under normoxia. This suggests that HIF2α plays an important role in regulating the organism-wide responses to oxygen availability.

Complement Factor H Related protein 5 (CFHR5) is a homologue of the complement regulating protein Complement Factor H and is of incompletely understood function. I uncovered an in-frame heterozygous duplication of exons 2 and 3 of the *CFHR5* gene which cosegregated with autosomal dominantly inherited microscopic and synpharyngitic macroscopic haematuria, glomerulonephritis and renal failure in 2 families with ancestry in the Troodos Mountains of Cyprus. The mutation results in the production of a protein with impaired affinity for complement deposited in the kidney. This disease, which I named CFHR5 nephropathy, is endemic in Cyprus, accounting for a significant proportion of renal disease on

the island, and may be amenable to systemic treatments. These findings implicate CFHR5 as a new and important regulator of complement in the human kidney.

Table of Contents

Table of figures

List of Abbreviations

aHUS	Atypical haemolytic uraemic syndrome
AMD	Age related macular degeneration
AP	Alternative pathway
ARMS	Amplification refractory mutagenesis system
ARNT	Aryl hydrocarbon nuclear translocator
ATP	Adenosine triphosphate
C3GN	C3 glomerulonephritis
cDNA	Complementary DNA
CFB	Complement factor B
CFH	Complement factor H
CFHR5	Complement factor H related protein 5
CFI	Complement factor I
CGH	Comparative genomic hybridization
CH	Conflicting homozygosity
cM	Centimorgan
CMS	Chronic mountain sickness
CNV	Copy number variation
CR	Complement receptor
CRP	C-reactive protein
DDD	Dense deposit disease
DFO	Desferrioxamine
ELISA	Enzyme linked immunosorbent assay
EPO	Erythropoietin
ESRD	End stage renal disease
FIH	Factor inhibiting HIF
FITC	Fluorescein isothiocyanate
GFR	Glomerular filtration rate
Hb	Haemoglobin
HIF	Hypoxia inducible factor
HRE	Hypoxia responsive element
IBD	Identical by descent
kDa	Kilodaltons
LCR	Low copy number repeat
LD	Linkage disequilibrium
LOD	Logarithm of the odds ratio
MAC	Membrane attack complex
MASP	MBL associated protease
MBL	Mannose binding lectin
MCGN	Mesangiocapillary glomerulonephritis
MLPA	Multiplex ligation dependent probe amplification
MRCA	Most recent common ancestor
mRNA	Messenger ribonucleic acid
NAHR	Non-allelic homologous recombination
N-TAD	N-terminal activation domain

ODDD	Oxygen dependent degradation domain
PCR	Polymerase chain reaction
PHD	Prolyl hydroxylase domain
PJS	Peutz Jeghers syndrome
PNH	Paroxysmal nocturnal haemoglobinuria
Pro	Proline
RCA	Regulators of complement activation gene cluster
SCR	Short consensus repeat domain
SNP	Single nucleotide polymorphism
STK11	Serine threonine kinase 11
VEGF	Vascular endothelial growth factor
VHL	von Hippel Lindau

Publications arising from this work

1. Daniel P. Gale, Sarah K. Harten, Cecil D. L. Reid, Edward G. D. Tuddenham and Patrick H. Maxwell. Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating HIF2{alpha} mutation. **Blood** 2008;112: 919-921

2. Daniel P Gale, Elena Goicoechea de Jorge, H Terence Cook, Rubén Martinez-Barricarte, Andreas Hadjisavvas, Adam G McLean, Charles D Pusey, Alkis Pierides, Kyriacos Kyriacou, Yiannis Athanasiou, Konstantinos Voskarides, Constantinos Deltas, Andrew Palmer, Véronique Frémeaux-Bacchi, Santiago Rodriguez de Cordoba, Patrick H Maxwell, Matthew C Pickering. Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis. **The Lancet** 2010;376(9743): 794 – 801.

Chapter 1: Introduction

1.1 The rationale for Genehunting

Elucidation of the molecular basis of disease has a number of advantages. Firstly, it can allow the development of a non-invasive (i.e. genetic) test which can aid in diagnosis; secondly, this can provide prognostic information for an individual or cohort; thirdly, understanding the molecular basis may itself provide insight into the pathophysiology of the disease and the physiology of the underlying biological pathway. In turn, this can pave the way towards rational development of treatment for that disease. Examples of this include the administration of recombinant alpha-galactosidase to people with Fabry's disease (Schiffmann, Kopp et al. 2001), and the use of the tyrosine kinase inhibitor Imatinib in chronic myeloid leukaemia (Druker, Talpaz et al. 2001).

In addition, understanding how a rare mutation can cause a rare disease may also provide information leading to the development of new treatments for more common diseases. An example of this is the appreciation that mutations in the *VHL* gene cause von Hippel Lindau disease which includes a predisposition to renal cell carcinoma. Subsequently, it was found that sporadic renal cell carcinoma, which accounts for approximately 3% of cancer deaths (or ~13,000 people in the USA) each year (Jemal, Siegel et al. 2009), in over 80% of cases is associated with somatic biallelic *VHL* inactivation (Barry and Krek 2004). The drugs Sorafenib and Sunitinib, which antagonise the receptor for vascular endothelial growth factor (which lies downstream of VHL, the product of the *VHL* gene), have shown efficacy in the treatment of sporadic renal cell carcinoma (Escudier, Eisen et al. 2007; Motzer, Hutson et al. 2007).

The ability to make a genetic diagnosis is particularly relevant in renal disease where decisions surrounding organ transplantation (depending for instance on whether the disease is likely to recur in an allograft or whether a particular individual is a suitable kidney donor) may be informed by the results of genetic testing.

It is proposed that successfully understanding an “experiment of nature,” where a change in a gene causes a change in the phenotype of a human being, can provide both clinical and academic benefits.

1.2 Historical perspective

Over the past few decades, rapid advances in molecular genetics have allowed elucidation of the molecular causes of a wide range of Mendelian disorders, including those causing some renal diseases, such as autosomal dominant polycystic kidney disease (European Polycystic Kidney Disease Consortium 1994). A key tool used to identify the genetic basis of high-penetrance Mendelian traits has been the linkage study, in which a set of genetic markers (or polymorphisms) are genotyped within a family or group of families. Within a family, those polymorphisms of which an allele is transmitted with the disease more often than would be expected by chance are likely to lie on a chromosomal segment near to the disease-causing variant and are considered to be linked with the disease. The likelihood that the observed pattern of cosegregation arose as a consequence of linkage divided by the likelihood that the observed pattern of cosegregation occurred by chance is known as the odds ratio, and its logarithm (conventionally in base 10) is termed the LOD score, which is the standard unit of linkage. Because of this logarithmic relationship, LOD scores of unrelated families sharing a phenotype (and hence assumed have mutations at the same locus) can be summated at a given chromosomal location, even if the markers and even the disease-causing mutation are different. Conventionally, since the prior probability of linkage for any chromosomal location is low (because the genome is so large), a LOD score of 3 (corresponding to an odds ratio of 1000:1 in favour of linkage underlying the observed pattern of recombinations) is regarded as significant. Clearly, the power of such studies (i.e. the maximum possible LOD score) is dependent on the number of observed (or inferred) transmissions of the disease and markers within the family (i.e. the number of informative meioses) and their resolution (i.e. the

precision with which they are able to define the boundaries of the disease-linked locus) is dependent both on the density of markers used and also on the number of (or rather distance between) the recombinations within the families.

Prior to the early years of the 21st century, genome-wide linkage studies were most commonly performed using up to several hundred polymorphic microsatellite markers distributed throughout the genome. Although each marker was likely to be informative, genotyping was slow and added significantly to the resources required for such studies. The cost of genotyping more markers across the genome did not justify the use of dense marker arrays and fine-mapping studies were performed on potentially linked regions to narrow down any linked interval as much as possible, with a fundamental limit being determined by the position of recombinations within the family (or families) under investigation. Analysis of the genotype data is not trivial and a range of computer programs were developed between the 1970s to the 1990s which allowed computation of LOD scores, both for single markers and for multiple markers (Ott 1974; Lathrop, Lalouel et al. 1984; Cottingham, Idury et al. 1993; O'Connell and Weeks 1995; Kruglyak, Daly et al. 1996; Gudbjartsson, Jonasson et al. 2000). These programs employ algorithms such as the Elston-Stewart algorithm (Elston and Stewart 1971) or the Lander-Green algorithm (Lander and Green 1987) in order to calculate linkage.

The Elston-Stewart algorithm uses a process known as peeling, where a large pedigree is split into smaller nuclear families. The LOD score is calculated for each nuclear family separately and these values summated across the whole pedigree. While efficient for large families, this approach requires all possible genotypes at each marker to be considered for each step, so with increasing numbers of markers the computational requirement increases exponentially. The Lander-Green algorithm works by expressing the flow of alleles as inheritance vectors according to whether an allele is on the maternal or paternal haplotype for each individual. For those markers where there is only one possible vector (i.e. the phase is

known) this is straightforward, however in many instances there will be a number of possible routes of transmission, defining a set of possible inheritance vectors. The likelihood for each marker is a function of the probability of each of the 2^{2i} (where *i* is the number of non-founders in the pedigree) inheritance vectors. The algorithm treats the state of each marker as a function only of the state of the adjacent markers (i.e. a Markov model[a]), based on the genetic distance between them. Consequently, increasing the number of markers causes a linear increase in the computational requirement. Since each inheritance vector contains 2^{2i} elements there is an exponentially increasing computational requirement with increasing family size. Some of these limitations can be overcome by using sampling techniques to test only a set of possible inheritance vectors in proportion to their likelihood (Sobel and Lange 1996) but in practice this type of solution also poses a significant challenge to modern computers, and recent publications have tended to result from other (more modern) techniques for the identification of genes responsible for human disease.

1.3 The whole genome era

The near-complete, annotated record of the human genome is now publicly available, and while the functions of many genes are still poorly understood, elucidation of their genomic location and their organism-wide expression distribution, combined with estimation of the predicted structure of their products and likely homologues has resulted in an exponential increase in what is understood about the molecular pathways responsible for health and disease. In addition, the growing body of experimental results available in public databases provides an expanding pool of knowledge which can allow researchers to hypothesize which genes are likely to play a role in contributing to a disease.

[a] A model obeying the Markov property is one in which the future state of the model depends only on the present state. An example is a frog hopping from lily-pad to lily-pad. If it cannot remember which lily-pad it was last on its choice of where to hop to depends only on its current location and therefore obeys the Markov property. Mathematical tools have been developed to describe the behaviour of such a system.

The identification of several million single nucleotide polymorphisms (SNPs) which are common across human populations has provided the substrate for the systematic analysis of genetic variation and its relationship to disease. The burgeoning field of genome science has benefitted, in the last decade, from technologies which have allowed rapidly expanding insight into the molecular genetic contribution to what are usually regarded as non-Mendelian, or sporadic, diseases – a field known as complex trait genetics. Central to the ability to genotype (i.e. determine which alleles are present in an individual) large numbers of SNPs is the Charge Coupled Device (CCD), a computer chip which is able to respond, with great 2-dimensional spatial resolution, to small changes in electromagnetic radiation (i.e. light) emitted from a chemical reaction. The CCD was invented over 40 years ago (Boyle and Smith 1970), but its development was catalyzed by the commercial imaging market, most notably the Sony Corporation which developed its use in handheld video recorders. Today the CCD has almost completely replaced photographic film in the consumer camera market, and Boyle and Smith received the 2006 Nobel Prize for Physics for their invention. The CCD, among other advances, has provided a crucial tool for modern SNP genotyping technology, allowing up to several million SNPs to be genotyped in an individual rapidly, using less than a microgram of genomic DNA and for a few hundred dollars.

These advances in high-throughput genotyping have allowed genetic variation among large cohorts of patients and controls to be studied, providing a new angle for the study of the molecular pathophysiology of common diseases (Wellcome Trust Case Control Consortium 2007). This type of study is powerful enough to detect even tiny risk effects of common variants but needs significant resources to gather DNA and phenotypic information from sufficient patients and controls drawn from a single population. Consequently, the genome wide association study (GWAS) approach has been less successful in identifying the genetic contribution to less common diseases, including many kidney diseases, where small numbers of patients allow detection only of alleles which confer a large amount of risk.

With growing understanding of the molecular structure of physiological systems and the pathways which contribute to disease, it is clear that the ability of researchers to identify candidate genes which might be responsible for a phenotype is increasing.

1.4 Genetic kidney disease

It is common for people with kidney failure not to receive a diagnosis, and this is particularly true for individuals who present over the age of 16. In the UK, the cause of renal failure in 15.8% patients starting renal replacement is categorised as 'aetiology uncertain' in the 20-65 years age group (Byrne, Ford et al. 2009), and although this percentage has fallen from 19.7% in 2004 (Ansell, Feest et al. 2004), there are clearly many young adults reaching end stage renal disease (ESRD) in whom the diagnosis is not known. It is notable that ESRD in childhood is less likely to be of unknown aetiology, with only 3% children requiring renal replacement therapy falling into this category in 2008 (Lewis 2008). The reasons for the failure to identify the cause of this significant contribution to renal disease are not clear. A contributing factor might be incomplete transfer of information to the registries, but nonetheless it is obvious that there are large numbers of patients receiving lifelong treatment for renal failure in the absence of an understanding of what their underlying disease is.

Epidemiological evidence that genetic factors might be important in the aetiology of some of this undiagnosed kidney disease has come from 2 different observations: Firstly, that risk of renal failure varies between different ethnic groups, even within the same geographical location – for instance South Asians have a higher risk of reaching ESRD than Caucasians residing in Canada (Barbour, Er et al. 2010). Secondly, there is marked familial clustering of renal disease in many different ethnic groups, with up to five-fold increased risk of renal failure seen in individuals with a first degree relative who has reached ESRD. Up to one third of African American patients receiving renal replacement therapy report a first or second degree relative

with ESRD, compared with 12-15% Caucasians (Satko, Sedor et al. 2007). Even when known Mendelian disorders (such as polycystic kidney disease and Alport syndrome) were excluded, 22.8% incident ESRD patients in one study reported a family history of ESRD (Freedman, Volkova et al. 2005). In this study, younger incident patients were more likely to have a family history than older ones, arguing against an environmental cause for the aggregation (since older people were hypothesised to have had greater exposure than the young to any putative environmental factors).

It therefore seems likely that unrecognised genetic factors are important in the aetiology of renal disease.

1.5 This study

The aim of this investigation was to use advances in genotyping technology combined with an integrated approach to molecular physiology to identify the molecular basis of disease in families with Mendelian inheritance of unusual phenotypes encountered in clinical practice. This utilised knowledge of biological pathways implicated by the phenotype, combined with SNP-based linkage analysis in small families to exclude as many candidate genes as possible and concentrate resequencing (and other techniques) on the best non-excluded candidate genes in order to identify a candidate mutation. Identification of a candidate mutation would be followed, where appropriate, by population-based, *in vitro* and clinical investigations to understand more about the disease.

Chapter 2: HIF2α Erythrocytosis and Pulmonary Hypertension

2.1 Introduction

The aim of this study was to understand the underlying cause of erythrocytosis (overproduction of red blood cells) which segregated as an autosomal dominant trait in a single family from North London who were originally described in 1979 (Howarth, Chanarin et al. 1979). Initial assessment had demonstrated that erythropoietin levels were inappropriately elevated in affected members of the family and, for reasons outlined below, this implied that the molecular defect was one involving altered oxygen sensing. This hypothesis was supported by the subsequent development of pulmonary arterial hypertension in older family members.

2.1.1 Oxygen sensing in eukaryotes

Oxygen is necessary for eukaryotic life and its predominant role is as the terminal electron acceptor for the respiratory chain, thus acting as the major substrate for energy metabolism. Complex multicellular animals, which are too large for all their cells to access oxygen by passive diffusion from the environment, have evolved respiratory and circulatory systems to provide oxygen to their organs and there is an oxygen gradient between the inspired air (20.9% oxygen with a pO_2 of 159 mmHg at sea level) and the tissues. In healthy humans, pO_2 is typically 25-60 mmHg in well vascularised organs, but much lower values of 8-10 mmHg are common in certain tissues including articular cartilage, renal medulla and epidermis (Vaupel, Kallinowski et al. 1989; Wang, Winlove et al. 2003; Cringle and Yu 2004; Gibson, Milner et al. 2008). Even within tissues, gradients of oxygenation exist, with subpopulations of the same cell-type experiencing different oxygen tensions within an organ. It has commonly been assumed that these variations in oxygenation are not important, and this assumption may have been based on the perspective that the level of oxygenation only becomes relevant at levels which are sufficiently low to compromise mitochondrial respiration, corresponding to a pO_2 of approximately 1 mmHg in the mitochondrion. Most studies have therefore concentrated on examining the

effects of compromised oxygen delivery where cessation of mitochondrial respiration leads to depletion of ATP and tissue necrosis.

It is now clear that sensing and responding to changes in oxygenation, even above the range at which mitochondrial function is limited by oxygen availability, is a critical attribute of the cells of complex organisms. Advances in the last 20 years have identified a highly conserved system, present in all eukaryotes which have been studied, based on the Hypoxia Inducible Factor (HIF) system (Kaelin and Ratcliffe 2008). This molecular apparatus responds to alterations in oxygen availability by coordinating a complex and pleiotropic transcriptional response which has effects at the cellular, local and organism-wide levels. Importantly, HIF is active within the physiological range of oxygen tension and has significant effects on gene expression and cellular behaviour at levels of oxygenation much higher than that at which mitochondrial respiration is compromised.

The molecular apparatus of the HIF pathway is now well understood, and the key insights which have unlocked its secrets have come both from experiments in model organisms and also clinical observations in humans.

2.1.2 Oxygen, haematocrit and erythropoietin

Evidence that oxygen plays a role in the regulation of haematocrit dates from the 19th century, when Bert and Jourdanet demonstrated that blood viscosity increased on exposure to altitude (Bert 1878) and Viault went on to show that high altitude exposure provokes a rise in blood haemoglobin concentration (Viault 1891). In 1906, Carnot and Deflandre injected serum from anaemic rabbits into normal rabbits and found that this led to an increase in haemoglobin concentration and proposed that this was mediated by a circulating factor (Carnot and Deflandre 1906). A key step was the development of reliable bioassays for erythropoietin (Cotes and Bangham 1961). Starting with the urine of patients with severe anaemia, Goldwasser succeeded in purifying erythropoietin (Miyake, Kung et al. 1977) leading to the

cloning of the EPO gene in 1985 (Jacobs, Shoemaker et al. 1985). Recombinant human erythropoietin was then developed to treat anaemia in patients with kidney failure and has been one of the most conspicuous success stories of molecular medicine (Winearls, Oliver et al. 1986; Winearls 1995). Isolation of the EPO gene also provided key tools for studying its regulation by hypoxia, including mRNA probes and adjacent regulatory sequences. As shown in ***Table 2.1-1***, conditions which lead to increased erythropoietin production and erythrocytosis are commonly associated with reduced tissue oxygen delivery. In contrast, uncoupling mitochondrial oxidative phosphorylation (for example with cyanide) does not increase erythropoietin expression (Necas and Thorling 1972; Tan and Ratcliffe 1991). This is because the underlying mechanism of oxygen sensing is not based on reduced ATP levels. Instead it is based on an enzyme reaction which uses molecular oxygen as a co-substrate. Cobalt increases erythropoietin production because at a molecular level it mimics the effect of low oxygen by disabling the "oxygen-sensing" enzyme reactions which regulate HIF (Goodman 1947; Semenza, Roth et al. 1994; Nytko, Spielmann et al. 2007). Some tumours which cause erythrocytosis are associated with genetic mutations in the *VHL* gene, which also lead to HIF activation (Wiesener, Seyfarth et al. 2002).

Conditions associated with hypoxaemia (reduced arterial pO_2)
Exposure to high altitude
Pulmonary diseases
Alveolar hypoventilation (e.g. owing to sleep apnoea or central nervous system disease)
Cardiovascular disease causing a right-to-left shunt (e.g. cyanotic congenital heart disease)
Conditions associated with normal arterial pO_2 but reduced oxygen delivery to tissues
Anaemia
Exposure to carbon monoxide
High affinity haemoglobin variants
Methaemoglobinaemia
2,3-Diphosphoglyceromutase deficiency
Conditions associated with normal oxygen delivery to tissues
Tumour-associated erythropoietin production
Renal disorders (e.g. polycystic kidney disease and Bartter's syndrome)
Cobalt toxicity
Chuvash polycythaemia
PHD2 erythrocytosis

Table 2.1-1 Conditions associated with increased erythropoietin production

2.1.2.1 Renal erythropoietin production

Organ ablation experiments in the 1950s first established that the kidney was essential for normal erythropoietic responses to low oxygen (Jacobson, Goldwasser et al. 1957). Subsequently it was shown that the gene for erythropoietin (the hormone responsible for mediating this response) is expressed at lower levels in a range of other cells and organs, including liver, brain, retina and testes (Eckardt, Pugh et al. 1994; Su, Wiltshire et al. 2004).

Under basal conditions, erythropoietin is expressed in the kidney by peritubular fibroblasts located near the boundary between cortex and medulla (Maxwell, Osmond et al. 1993). When oxygen delivery to the kidney is reduced, fibroblasts located farther from the cortico-medullary

junction are progressively recruited to produce erythropoietin. There is some evidence that each fibroblast is recruited in an all-or-none manner (Eckardt, Koury et al. 1993).

The signal for respiratory control detected by the carotid body is arterial pO_2. This would not be suitable for regulating erythropoietin production since it is not dependent on haemoglobin concentration. Instead the system must measure the amount of oxygen delivered by the blood. Illustrating this, carbon monoxide exposure does not reduce arterial pO_2, but decreases the blood oxygen content leading to increased erythropoietin production. This also explains why high affinity haemoglobin variants are associated with erythrocytosis. The physiology is incompletely understood but is based on sensing *tissue* pO_2.

At least two aspects of the kidney probably make it suitable for this task. First, the kidneys receive a high blood flow – although they account for less than 0.5% of the body mass in humans they receive approximately 20% of the cardiac output. This enormous blood flow is maintained by complex neuroendocrine autoregulatory mechanisms, so that renal blood flow is rather stable – even in the face of quite marked haemodynamic disturbance. Second, oxygen consumption in the kidneys mainly relates to resorption of sodium in the tubules, which is dependent on glomerular filtration rate (GFR). Any reduction in renal blood supply owing to the autoregulatory mechanisms being overwhelmed is accompanied by a fall in GFR, reduced sodium resorption, and hence decreased local oxygen consumption. Consequently, renal artery stenosis does not generally cause erythrocytosis. Supporting the model, experimental administration of acetazolamide, which reduces sodium reuptake and hence oxygen consumption inhibits the erythropoietin response to hypoxia (Eckardt, Kurtz et al. 1989).

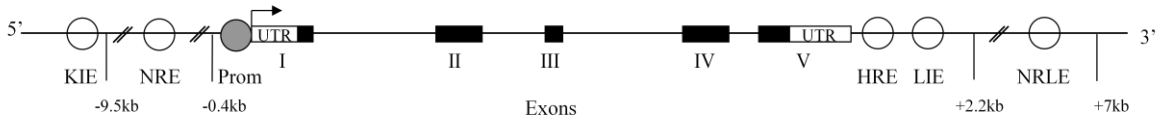
2.1.2.2 Hepatic Erythropoietin production

Although the kidneys are the primary source of erythropoietin in adult mammals, an appreciable amount of erythropoietin is produced in the liver by hepatocytes and interstitial stellate cells – also known as Ito cells (Koury, Bondurant et al. 1991; Maxwell, Ferguson et al.

1994). Hepatic erythropoietin production is regulated by oxygen tension, with the least oxygenated part of the liver lobules responsible for most erythropoietin production. Curiously, cultured renal fibroblasts do not produce erythropoietin, although the reasons for this are not known. Fortunately, certain human hepatoma cell lines have provided an excellent model system for studies of the regulation of EPO expression by oxygen (Goldberg, Glass et al. 1987).

2.1.3 Regulation of erythropoietin gene expression

Three non-coding regions of the EPO gene are highly conserved among mammalian species – the promoter region, the first intron and an enhancer region lying approximately 100bp 3' to exon 5. Mapping experiments using mice with transgenes containing mouse or human erythropoietin sequence have demonstrated the presence and location of elements concerned with tissue-specific gene expression; some of these elements are located many kilobases 5' to the gene (***Figure 2.1-1***). A negative regulatory element located closer to the gene is important for suppressing expression in most tissues (Semenza, Dureza et al. 1990).



KIE	Kidney inducible element
NRE	Negative regulatory element
Prom	Promoter
HRE	Hypoxia responsive element
LIE	Liver inducible element
NRLE	Negative regulatory liver element
UTR	Untranslated Region

Figure 2.1-1 The structure of the EPO gene

The relative positions of the exons, promoter and enhancer elements responsible for the normal regulation of the human *EPO* gene (not to scale). Binding of HIF and other transcriptional co-activators to HRE is responsible for oxygen-dependent activation of *EPO* transcription. Transcription factors binding to the other enhancer/repressor elements determine organ-specific regulation of *EPO* gene expression.

Experiments using human hepatoma cells identified an enhancer element downstream of the polyadenylation site which confers hypoxia responsiveness and is termed an Hypoxia Responsive Element (HRE) (Imagawa, Goldberg et al. 1991; Pugh, Tan et al. 1991; Madan and

Curtin 1993). The identification of the HRE allowed isolation of a protein complex that binds to it – the Hypoxia Inducible Factor (HIF) complex (Wang and Semenza 1993). Subsequently, it has been shown that numerous other genes contain an HRE in their flanking sequence, with the core consensus sequence (RCGTG). Further, all mammalian cells – not just those that produce erythropoietin - activate HIF in hypoxia (Beck, Weinmann et al. 1993; Maxwell, Pugh et al. 1993; Wang and Semenza 1993) and that a large number of genes, in addition to EPO have an HRE and are upregulated by HIF activation (O'Rourke, Dachs et al. 1997). Thus HIF is a widely operative oxygen-responsive transcription control system, and it regulates many other responses besides increased erythropoiesis.

2.1.4 Oxygen sensing by HIF

HIF is a heterodimeric transcription factor composed of alpha and beta subunits, termed HIF-α and HIF-β respectively. The HIF-α subunit is oxygen-labile and unique to the HIF pathway, whereas the β subunit is constitutive and also heterodimerises with other transcription factors. It is also known as aryl hydrocarbon receptor nuclear translocator (ARNT). There are at least three genes for HIF-α, termed *HIF1*α, *-2*α and *-3*α respectively, with additional complexity arising from the production of alternatively spliced transcripts. A comparison of HIF1α and HIF2α is shown in ***Figure 2.1-2***. In hypoxic conditions, the alpha subunits are able to translocate to the nucleus and dimerize with the beta subunits. These bind to HREs, recruit co-activators and increase transcription of target genes. In addition, HIF can modulate gene expression by binding to other transcription factors.

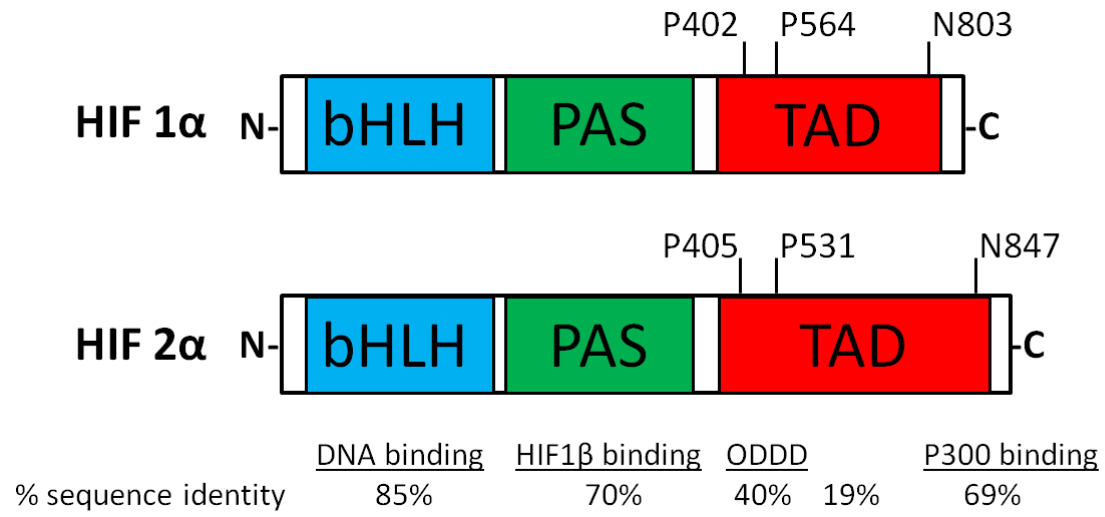


Figure 2.1-2 The Structure of HIF-1α and HIF-2α

Amino acid similarity (%) is shown between the two proteins at each domain: DNA-binding (blue); HIF1β-binding (green); Terminal activation domain (red), including the oxygen dependent degradation domain (ODDD) and C-terminal activation/P300 binding domain. The residues involved in enzymatic hydroxylation are also shown above each molecule. Adapted from (Hu, Sataur et al. 2007).

2.1.5 Regulation of HIF by oxygen

In the presence of oxygen, the HIF-α subunit is hydroxylated at one of two conserved proline residues (Pro-402 and -564 of human HIF1α and Pro-405 and -531 of HIF2α) a reaction catalysed by Prolyl Hydroxylase Domain (PHD) enzymes. This prolyl-hydroxylation allows interaction with the von Hippel-Lindau E3 ubiquitin ligase (VHL) which leads to ubiquitin-mediated targeting of HIF-α for degradation by the proteasome (Maxwell, Wiesener et al. 1999). In addition, hydroxylation of a conserved asparagine residue (Asn-803 of human HIF1α and Asn-851 of HIF2α) in the C-terminal transactivation domain occurs in the presence of oxygen (see ***Figure 2.1-3***). This asparaginyl hydroxylation prevents recruitment to the HIF complex of co-activators (Hewitson, McNeill et al. 2002; Lando, Peet et al. 2002).

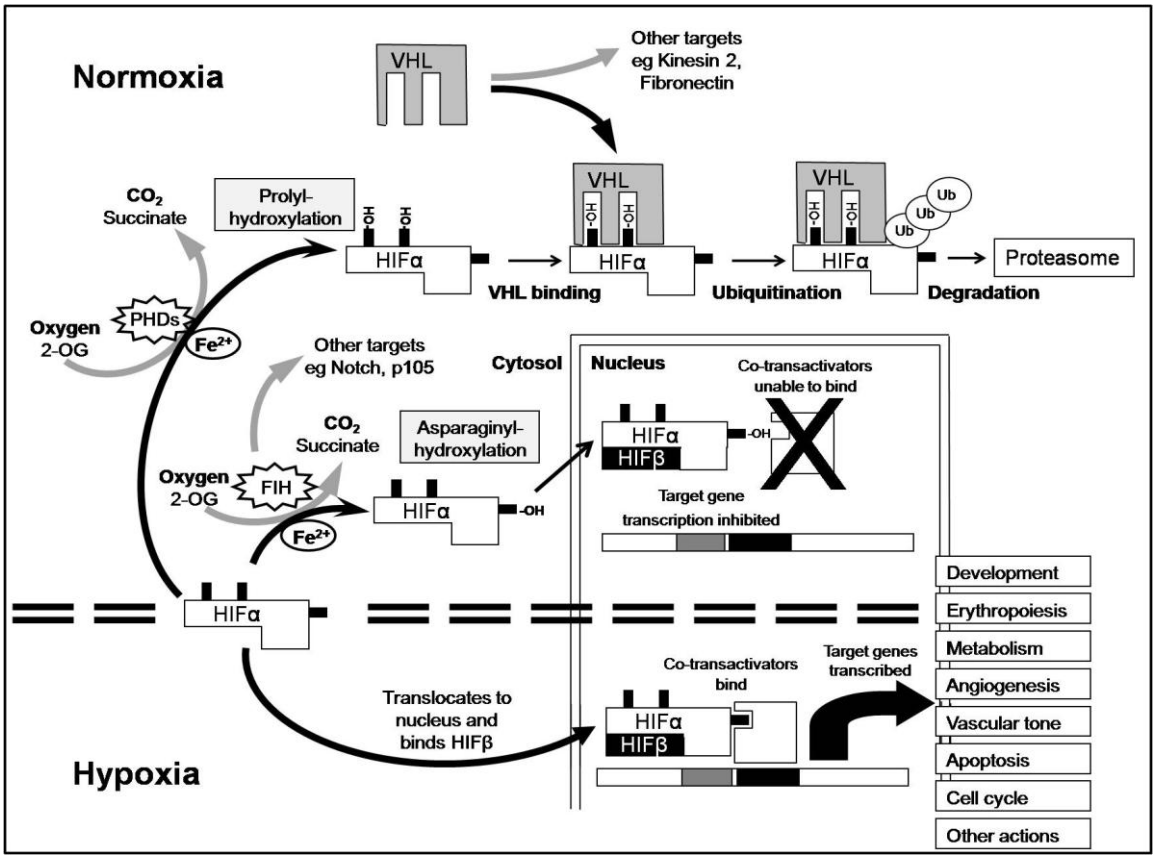


Figure 2.1-3 Regulation of Hypoxia Inducible Factor (HIF)

In the presence of oxygen HIF-α undergoes prolyl hydroxylation at conserved residues, catalysed by a family of iron(II) (Fe^{2+}) and 2-oxoglutarate (2-OG) dependent prolyl hydroxylase domain (PHD) enzymes leading to recognition by the von Hippel Lindau (VHL) protein which catalyses ubiquitination, targeting HIF-α for proteasomal degradation. In addition, a conserved asparagine residue undergoes hydroxylation, by Factor Inhibiting HIF (FIH), which blocks activation of HIF target genes, as well as having other functions. In hypoxic conditions, HIF-α translocates to the nucleus, binds to HIF-β and additional co-transactivators to activate transcription of target genes which, in turn, modulate a wide variety of cellular processes.

2.1.5.1 Regulation of HIF destruction by PHD enzymes

Prolyl-hydroxylation of HIF-α subunits occurs through the action of prolyl-hydroxylase domain (PHD) enzymes of which there are at least three in humans, termed PHD1, 2 and 3 (Epstein, Gleadle et al. 2001). These enzymes are non-haem iron(II)- and 2-oxoglutarate dependent dioxygenases which have homologues in all eukaryotes and even some prokaryotes (McDonough, Li et al. 2006). In the presence of molecular oxygen they produce HIF-α–OH, CO_2 and succinate (***Figure 2.1-4***).

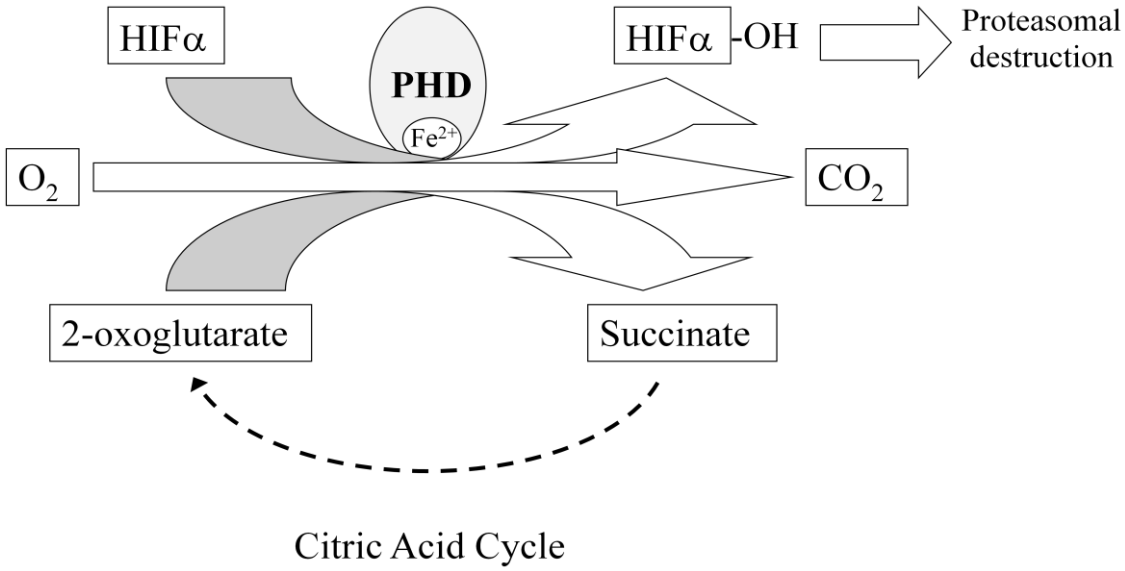


Figure 2.1-4 Prolyl Hydroxylase Domain (PHD) enzymes

PHD enzymes require O_2 and 2-oxoglutarate as co-substrates and non-haem ferrous iron (Fe^{2+}) as a co-factor to hydroxylate conserved proline residues of HIF-α, allowing recognition of HIF-α-OH by VHL leading to its subsequent ubiquitination and degradation. The particularly low affinity of PHD enzymes for oxygen makes molecular oxygen concentration the determinant of enzymatic activity in most physiological conditions. However, significant depletion of 2-oxoglutarate or accumulation of succinate (for instance by deficiency of citric acid cycle enzymes or co-substrates) may inhibit PHD enzymatic activity. Furthermore, deficiency of ferrous iron (Fe^{2+}), for instance by excessive reactive oxygen species production, may also inhibit PHD activity and allow un-hydroxylated HIF-α to accumulate even when molecular oxygen is plentiful.

The PHD enzymes are clearly suitable to act as a cellular "oxygen sensor". Importantly, the availability of the other co-substrate (2-oxoglutarate) and of Fe^{2+} (which is present at the active site) could also affect PHD activity. Cobalt(II), by substituting for iron(II), disables PHD enzymes – explaining why it activates HIF in cell culture and causes erythrocytosis when administered to mammals (Goodman 1947; Nytko, Spielmann et al. 2007).

2-oxoglutarate is an intermediate in the citric acid cycle. Other citric acid cycle intermediates, succinate and fumarate, competitively inhibit the interaction of 2-oxoglutarate with the PHD enzymes *in vitro*. *In vivo*, this may result in an important cross-talk between metabolism and HIF activation. Supporting this possibility, genetic defects in succinate

dehydrogenase and fumarate hydratase are associated with evidence of increased HIF activity (Pollard, Briere et al. 2005; Pollard, Spencer-Dene et al. 2007; Ratcliffe 2007).

The PHD enzymes are highly conserved and were initially identified by their homology to the oxygen-dependent dioxygenase *Egl9* in the nematode *C. elegans* (Epstein, Gleadle et al. 2001). It is noteworthy that the molecular oxygen-sensing apparatus is conserved between mammals and organisms lacking EPO, blood vessels, or a haematopoietic system.

The differential roles of the three PHD enzymes are not yet fully understood, but it is likely that their differing affinity for the two proline target residues in HIF-α is important. Since the HIF hydroxylation reaction is not at equilibrium the amount of PHD enzyme present influences the level of HIF activity. Interestingly, expression of the PHD isoforms in rats is variable from one tissue to another, with PHD2 being ubiquitously expressed, and PHD1 and PHD3 being most abundant in the testis and heart respectively. While PHD1 appeared to be constitutively expressed, PHD2 and 3 were inducible by hypoxic exposure, possibly representing a feedback loop (Willam, Maxwell et al. 2006). At a subcellular level, PHD1 is predominantly expressed in the nucleus, PHD2 in the cytoplasm and PHD3 in both compartments (Metzen, Berchner-Pfannschmidt et al. 2003).

siRNA knockdown experiments in cell lines have shown that altered expression of all three PHD enzymes influences HIF activation in specific contexts, and that in most cell types PHD2 is the predominant isoform. Consistent with this, absence of PHD1 or PHD3 does not have obvious consequences in knockout mice, but PHD2 is required for embryonic development (Takeda, Cowan et al. 2007). PHD2 appears to have a key role in EPO regulation in humans, since heterozygosity for a PHD2 missense mutation is associated with autosomal dominant erythrocytosis (see below).

2.1.5.2 Factor Inhibiting HIF

Like the PHD enzymes, Factor Inhibiting HIF (FIH) is a 2-oxoglutarate and iron(II)-dependent dioxygenase. In the presence of oxygen, FIH hydroxylates a conserved asparaginyl residue in the C-terminal transactivation domain of HIF-α, causing repression of HIF mediated transcription by inhibiting the recruitment of co-activators. Under *in vitro* conditions the K_m of the PHD enzymes for oxygen is in the range 230-250 µmol/l (water saturated with air has an oxygen concentration of approximately 200 µmol/l), whereas the K_m of FIH for oxygen is 90 µmol/l (Hirsila, Koivunen et al. 2003; Koivunen, Hirsila et al. 2004). Consistent with this, functional experiments suggest that FIH is more important in regulating HIF under severe hypoxic conditions in cell lines (Stolze, Tian et al. 2004).

As well as being dependent on the availability of molecular oxygen, all four HIF hydroxylase enzymes require iron(II) and 2-oxoglutarate to function and are sensitive to highly oxidising environments (possibly owing to the conversion of iron(II) to iron(III)), so that the availability of reducing agents (such as ascorbate) can exert a significant influence on HIF activity (Pan, Mansfield et al. 2007; Vissers, Gunningham et al. 2007; Vissers and Wilkie 2007).

2.1.6 Effects of HIF activation

HIF activation alters the expression of many genes, and is activated in the physiological range of oxygen tension. Physiological processes dependent on HIF target genes include erythropoiesis, iron transport, angiogenesis and development, vascular tone, cellular energy metabolism, matrix metabolism, apoptosis and others (Schofield and Ratcliffe 2004). Given this pleiotropic range of actions, the central role of oxygen in cellular metabolism and the fact that oxygen delivery to tissue is reduced in many pathological settings it is not surprising that the HIF pathway plays a significant role in mammalian development, physiology and pathology.

To date, the disease setting in which HIF is most strikingly involved is cancer. This illustrates the consequences of mutations in the pathway in humans and the potential of the pathway for

regulating angiogenesis and cellular energy metabolism (Maxwell, Dachs et al. 1997; Gordan and Simon 2007). Inheriting one defective *VHL* allele results in von Hippel Lindau disease, a multisystem disorder characterised by a very high risk of haemangioblastomas in the eye and central nervous system, renal cell carcinoma, phaeochromocytoma and other tumours (Kaelin 2008). The tumours are associated with somatic mutation or inactivation of the remaining VHL allele and constitutive HIF activation. Sporadic renal cell carcinomas are also associated with biallelic VHL loss-of-function in the great majority of cases (Gnarra, Tory et al. 1994) and re-expressing VHL in cells derived from these tumours efficiently suppresses growth of experimental tumours (Zimmer, Doucette et al. 2004). Although VHL has other roles besides regulating HIF, activation of HIF appears to be the pivotal consequence of VHL loss-of-function, at least in the kidney.

Apart from the role of genetic activation of HIF in renal cancer, activation of HIF through local hypoxia in other cancers is common (Serganova, Humm et al. 2006). In addition, other signals which are commonly activated in cancer cells interact with the pathway, and are able to provoke a HIF-mediated response even when oxygen levels are not reduced, or potentiate the HIF response at a particular level of oxygen (Laughner, Taghavi et al. 2001; Alam, Maizels et al. 2004; Zhou, Schmid et al. 2004; Treins, Giorgetti-Peraldi et al. 2005; Peng, Karna et al. 2006). Key consequences of HIF activation in cancer are the effect on signalling from cancer cells to the vasculature, and on metabolism in the cancer cell. As tumours grow they require angiogenesis to provide oxygen (Folkman, Watson et al. 1989) and the role of HIF in this process is now well recognised. A number of angiogenic factors, including vascular endothelial growth factor (VEGF), platelet derived growth factor (PDGF), angiopoietin-1 and -2, Flt-1 (a VEGF receptor) are HIF targets, and their upregulation by HIF is consistent with a cellular adaptation to improve blood supply in hypoxia (Pugh and Ratcliffe 2003). Furthermore, HIF targets such as inducible nitric oxide synthase (iNOS), PDGF-β receptor and cyclooxygenase

(COX)-2 can also modify the behaviour of vascular smooth muscle cells to improve perfusion in hypoxic regions (Fong 2008).

Another striking feature of activation of HIF is that it results in a comprehensive reprogramming of energy metabolism. First, it increases expression of glucose transporters and glycolytic enzymes, resulting in increased glucose uptake, glycolysis and lactate production (Semenza, Roth et al. 1994; Ebert, Firth et al. 1995; Firth, Ebert et al. 1995). Second, by increasing expression of pyruvate dehydrogenase kinase, it decreases entry of pyruvate into the citric acid cycle (Kim, Tchernyshyov et al. 2006; Zhang, Gao et al. 2007). Third, it alters expression of components of cytochrome c oxidase expression (Fukuda, Zhang et al. 2007). This major regulation of cellular energy metabolism is a key function of the HIF system and has important implications for cells, such as those of the immune system, which must adapt to a range of oxygen tensions.

In addition to studies on metabolism in cancer cells, it is now becoming clear that HIF activation can have a powerful influence on energy utilization in other cell types. For example, studies of haematopoietic cells have demonstrated that growth factor stimulation is necessary for HIF1α expression and that, when growing, HIF1α is necessary for these cells to survive in hypoxia. This is thought to result, at least in part, from the ability of growth factor-induced HIF1α to reprogram cellular glucose utilization from anabolic synthesis to lactate (i.e. energy) production, improving survival at the expense of growth when a key substrate (oxygen) becomes limiting (Lum, Bui et al. 2007).

The molecular mechanisms underlying the cross-talk between growth factor signalling and cellular oxygen sensing are complex and only partly understood, but studies both in cancer cells and using other approaches, have established that HIF has complex bidirectional interactions with other signalling systems such as the PI3 kinase/AKT/mTOR and NFκB pathways (for review see (Gale and Maxwell 2010)).

2.1.7 Role of different HIF subunits in the EPO response

HIF can activate the transcription of a large number of genes (***Figure 2.1-5***). A major challenge is to understand how the hypoxic response of different cell types is selected from this repertoire. Some selectivity probably arises from the different HIF-α subunits each of which is encoded a different gene.

Recently, it has been appreciated that HIF1α is very widely expressed and appears to have more selectivity for regulating changes in energy metabolism (Papandreou, Cairns et al. 2006; Pan, Mansfield et al. 2007; Webb, Coleman et al. 2009). Conversely, HIF2α expression is restricted both during development and in adult animals and appears to correlate with the proliferative phenotype of hypoxic and HIF-dysregulated cells and organisms (Sowter, Raval et al. 2003; Gordan, Lal et al. 2008; Webb, Coleman et al. 2009). While there are differences in the structure of the subunits, the significance of these differences is not yet well understood (see ***Figure 2.1-2***). In a construct-reporter gene system it was found that the N-terminal activation domain (N-TAD) was responsible for conferring HIF1α or HIF2α target gene specificity whereas the C-TAD promoted the expression of target genes common to both subunits (Hu, Sataur et al. 2007). Less is known concerning HIF3α which encodes several alternative splice forms, at least one of which can function as a dominant negative regulator of HIF (Makino, Cao et al. 2001; Maynard, Evans et al. 2007). Which of the HIF-α subunits is predominantly responsible for the regulation of erythropoietin is not known, with evidence existing for both HIF1α and HIF2α playing a role (Yeo, Cho et al. 2008).

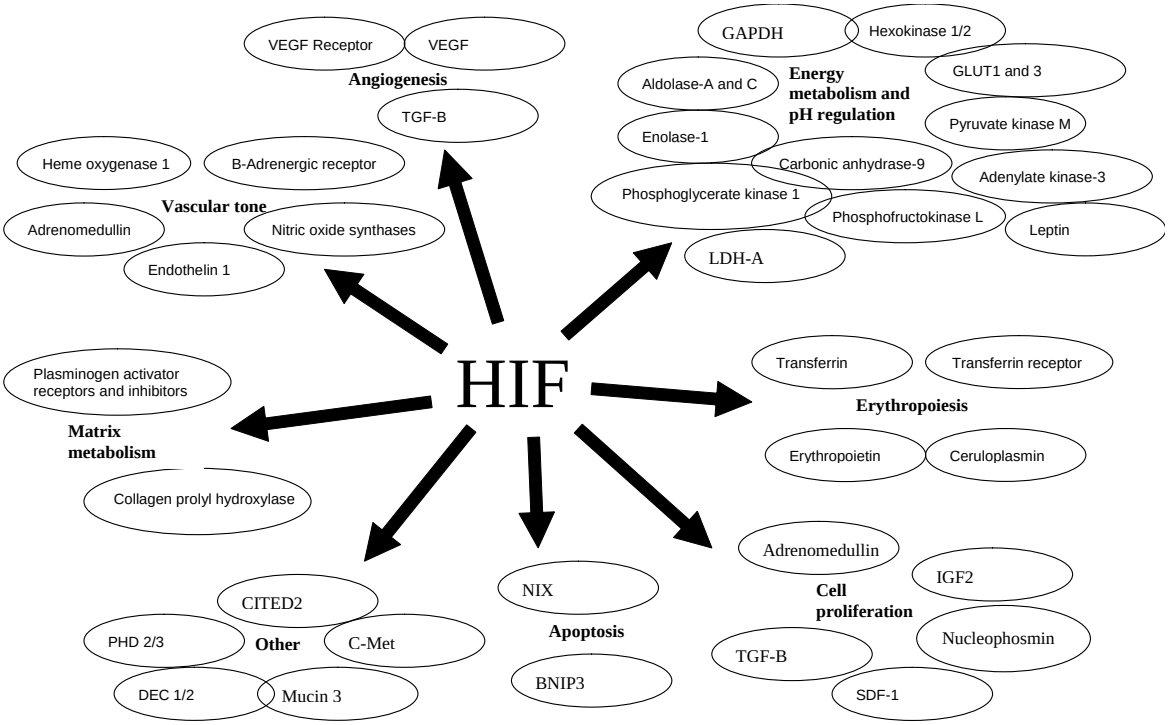


Figure 2.1-5 Selected HIF target genes

Studies in genetically modified mice have also provided evidence that HIF1α and HIF2α have distinct roles. HIF1α and HIF2α are clearly non-redundant in mice, since knocking out either gene has major consequences for embryonic development (Webb, Coleman et al. 2009). HIF2α is particularly important in regulating EPO production, and inactivation of *hif2α* in mice results in reduced EPO expression and anaemia which is reversed by administration of exogenous erythropoietin (Scortegagna, Ding et al. 2005; Gruber, Hu et al. 2007). HIF2α knockdown in the mouse retina also reduces local EPO production strikingly (Morita, Ohneda et al. 2003; Ding, Scortegagna et al. 2005). Consistent with a dominant role for HIF2α in EPO regulation, in hypoxic rat liver and kidney, the hepatocytes and renal cortical fibroblasts express HIF2α, but little or no HIF1α (Rosenberger, Mandriota et al. 2002). Nevertheless, HIF1α likely also contributes to regulating EPO production, since the EPO response to hypoxia is reduced in hif1α +/- mice (Yu, Shimoda et al. 1999).

2.1.8 Genetic defects in the oxygen sensing pathway in humans

Insights into how biological responses to hypoxia are controlled have come from disorders in which oxygen sensing is altered. To date, three distinct genetic defects of oxygen sensing have been recognised which have given insights into the role of the HIF system in mammalian oxygen homeostasis.

2.1.8.1 Von Hippel-Lindau disease

Von Hippel-Lindau disease is an autosomal dominant condition characterised by highly vascular tumours, most notably clear cell renal cell carcinoma, phaeochromocytoma and haemangioblastomas of the central nervous system and retina. The *VHL* gene was identified by positional cloning in 1993. VHL (the protein product of *VHL*) plays a crucial role in the degradation of prolyl-hydroxylated HIF-α by the proteasome (Maxwell, Wiesener et al. 1999). One defective *VHL* allele is not sufficient to activate HIF, but somatic inactivation of the second allele in susceptible tissue leads to HIF activation, and following other events, to the development of a tumour. About 70% of sporadic clear cell renal cell carcinomas have defects in VHL, and re-expression of VHL suppresses tumour growth in xenograft assays.

A hallmark of VHL defective tumours is HIF activation, and the biological role of HIF target genes such as vascular endothelial growth factor in angiogenesis is consistent with the highly vascular nature of the tumours of VHL disease. There is good evidence that control of HIF is a key aspect of VHL's tumour suppressor actions, although other pathways almost certainly contribute (Wykoff, Pugh et al. 2000).

Interestingly, neutrophils in VHL patients show a partial hypoxic phenotype showing that in these cells one functional copy of the VHL gene is insufficient to regulate HIF normally (Walmsley, Cowburn et al. 2006). But patients with VHL disease do not have erythrocytosis (unless there is ectopic production by a renal cell carcinoma) implying that a single normal *VHL* allele is enough to result in suppressed erythropoietin production in normoxia.

2.1.8.2 Chuvash Polycythaemia

In 1974 endemic hereditary erythrocytosis was reported in the population of Chuvashia, in the mid-Volga region of Russia (Poliakova 1974). Affected individuals are homozygous for a C→T mutation at nucleotide 598 of the *VHL* coding sequence which results in an arginine to tryptophan substitution at amino acid 200 (R200W) of VHL (Ang, Chen et al. 2002; Ang, Chen et al. 2002). This change impairs the interaction of VHL with hydroxylated HIF-α causing relative overactivity of HIF. Haplotype analysis has indicated that this mutation is due to a founder effect originating 14,000-62,000 years ago (Liu, Ang et al. 2004).

Chuvash Polycythaemia is characterised by elevated haemoglobin (mean haemoglobin in untreated affected cases is 22.6 +/- 1.4 g/dl) with inappropriately elevated levels of EPO. Other haematological parameters are normal. There is an increased incidence of vertebral haemangiomas, varicose veins and thrombosis. Premature mortality is increased, largely as a result of cerebrovascular disease and heart failure and, untreated, the life expectancy is 42 years. Intriguingly, the response of erythrocyte precursors to erythropoietin *ex vivo* is exaggerated in affected individuals, which is consistent with the view that HIF regulates not only erythropoietin expression but also sensitivity, although how this is mediated is not clear (Ang, Chen et al. 2002).

Unlike VHL disease, increased risk of malignancy is not a feature of Chuvash Polycythaemia. This may be because the mutated VHL is active enough to suppress HIF activity sufficient to prevent tumour genesis. Alternatively, it is possible that the mutated VHL has preserved affinity for prolyl-hydroxylated HIF2α sufficient to ensure that the proliferative phenotype of hypoxic exposure remains suppressed.

Individuals with Chuvash Polycythaemia offer an incisive “experiment of nature” in which VHL function is reduced. Potentially they can give insight into what other aspects of physiology are regulated by VHL. Affected individuals have mildly elevated pulmonary arterial blood

pressure and an exaggerated pulmonary arterial vasoconstriction response to mild hypoxic exposure (Smith, Brooks et al. 2006) suggesting that VHL has a role in these aspects of human physiology. Interestingly, systemic resting blood pressure is lower than in control subjects (Smith, Brooks et al. 2006). In addition, average resting pulmonary arterial pressures were elevated in a cohort of patients with Chuvash patients compared with controls (Bushuev, Miasnikova et al. 2006). More recently, it was shown that transgenic mice homozygous for the same R200W mutation in *vhl* develop pulmonary hypertension in addition to erythrocytosis (Hickey, Richardson et al. 2010).

2.1.8.3 PHD2 erythrocytosis

Autosomal dominant inheritance due to a C→G change at nucleotide 950 in *PHD2* has been described in a single UK family (Percy, Zhao et al. 2006). The clinical features include relatively mild erythrocytosis with inappropriately non-suppressed erythropoietin levels. The mutation predicts a substitution of arginine for proline at codon 317 of the protein – a residue which is close to two of the residues coordinating the ferrous atom at the active site of the enzyme. Mutant PHD2 with this substitution has a reduced ability to bind to HIF1α and HIF2α and reduced prolyl-hydroxylase activity, and the disease mechanism is thought to be haploinsufficiency. More recently, further mutations in PHD2 have been implicated in erythrocytosis with similar clinical features (Percy, Furlow et al. 2007; Al-Sheikh, Moradkhani et al. 2008).

The observation that neither Chuvash polycythaemia nor PHD2 erythrocytosis are associated with an obvious increased risk of renal cell carcinoma illustrates the point that the ability of a cell to degrade HIF in oxygen must be almost completely ablated in order to cause tumorigenesis, or alternatively that some other function of VHL needs to be disabled. At a cellular level, VHL disease is recessive since it involves biallelic loss of VHL function through a somatic “second hit”.

2.1.8.4 A novel HIF pathway disease?

A family has been identified in North London in whom erythrocytosis with raised circulating erythropoietin levels is transmitted in an autosomal dominant fashion and in whom exonic PCR amplification and sequencing has revealed no mutation in *PHD2* or *VHL*. Two older affected members of this pedigree were recently diagnosed with pulmonary arterial hypertension and it seemed possible that a defect in HIF regulation is responsible for all the aspects of the phenotype. The aim of this investigation was to identify the manifestations of the disease and to understand its genetic basis.

2.2 Clinical histories

The family was first described in 1979 (Howarth, Chanarin et al. 1979). The index case (individual II-1 in ***Figure 2.2-1***) had been diagnosed with erythrocytosis at the age of 27 following attendance at an antenatal clinic with her first pregnancy. She was entirely fit and well and her presenting haemoglobin (Hb) concentration was 18.1 g/dl. She was treated with 500ml venesection (which reduced the Hb to 16.8 g/dl) and proceeded to uneventful labour and delivery of a healthy girl (designated III-2).

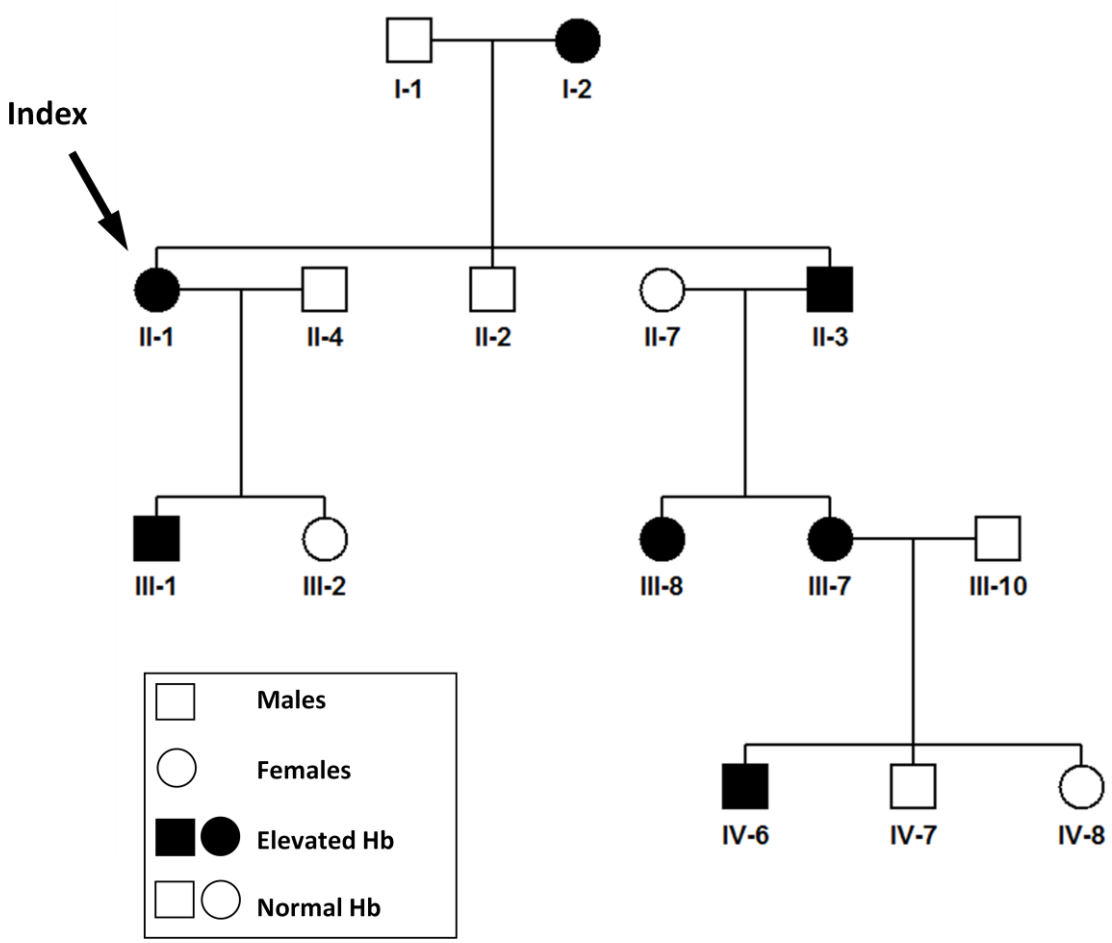


Figure 2.2-1 Erythrocytosis pedigree
Not standard nomenclature – additional unaffected relatives not shown

4 months post-partum, the Hb had risen to 20.2g/dl (haematocrit 60.8%) and clinical examination was normal with no splenomegaly identified clinically or by technetium scanning.

The mean cell volume was 92 fl with a normal mean cell Hb concentration. Red cell mass was 36ml/kg (196% predicted value corrected for body surface area) and plasma volume 28 ml/kg (normal range 40-50 ml/kg). The arterial pO_2 was 13.2 kPa while breathing room air and other investigations, including arterial oxygen saturation, O_2 dissociation curves and intravenous pyelogram were normal. Serum biochemistry (including lactate dehydrogenase and iron studies) was normal, as was peripheral differential blood cell count (***Table 2.2-1***). Bone marrow aspiration showed normoblastic erythropoiesis and iron was detectable (***Figure 2.2-2***). She had smoked 10 cigarettes a day since the age of 21. Treatment with recurrent venesection was administered and recovery of the Hb to >20g/dl in 6 months was observed if venesection was not performed. Between 1979 and 1989 she was also treated with hydroxyurea and a short course of chlorambucil. In 1989 the serum erythropoietin level was 1637 mIU/ml (NR 9.1-23 mIU/ml) after venesection to an Hb of 15 g/dl and was 361 mIU/ml immediately prior to venesection on a subsequent occasion some years later. Erythropoietin levels subsequently rose to >2000 by the year 2000 at which time the haematocrit was normal with venesection sufficient to maintain marked iron deficiency with a mean cell volume <60 fl.

Parameter	**Normal Range (adult)**	**II-1**	**II-3**	**I-2**	**III-7**	**III-8**	**III-1**	**IV-6**
Hemoglobin (g/dL)	*11.5-17.5*	20.2*	18.9*	22.8	19.7	18.6	20.2	16.1†
EPO (mIU/mL)	*9.1-23*	>2000*	1980*	–	26.2	28.1	13.9	–
Total red cell/plasma volume (mL/kg)	*22-32/ 40-50*	33/28	56.6/–	34/22.4	–	–	–	–
Pulmonary Hypertension		Present	Present	–	–	–	–	–

Table 2.2-1 Laboratory parameters
*Not contemporaneous measurements
†Sampled at 5 years of age - mean value in this age group is 12.5 g/dl

In 2002, at the age of 54, she was admitted to hospital with breathlessness and ankle swelling. Oxygen saturation was 92-94% on room air. Ventilation-perfusion scanning and subsequent computer tomographic pulmonary angiography revealed no evidence of pulmonary embolic disease and echocardiography revealed significant pulmonary arterial hypertension (estimated at 90mmHg) which was confirmed on subsequent cardiac catheterisation (***Table 2.2-2***). Left ventricular function was normal with no significant valvular pathology. Treatment with warfarin was started, however her symptoms of breathlessness worsened and she stopped work as a cleaner in 2004. She stopped smoking in 2005 following an episode of atrial flutter, treated with ablation and insertion of a permanent pacemaker. Treatment with bosentan (an endothelin receptor antagonist) provided some symptomatic relief however her oxygenation and mobility gradually decreased and she died in 2006 aged 58.

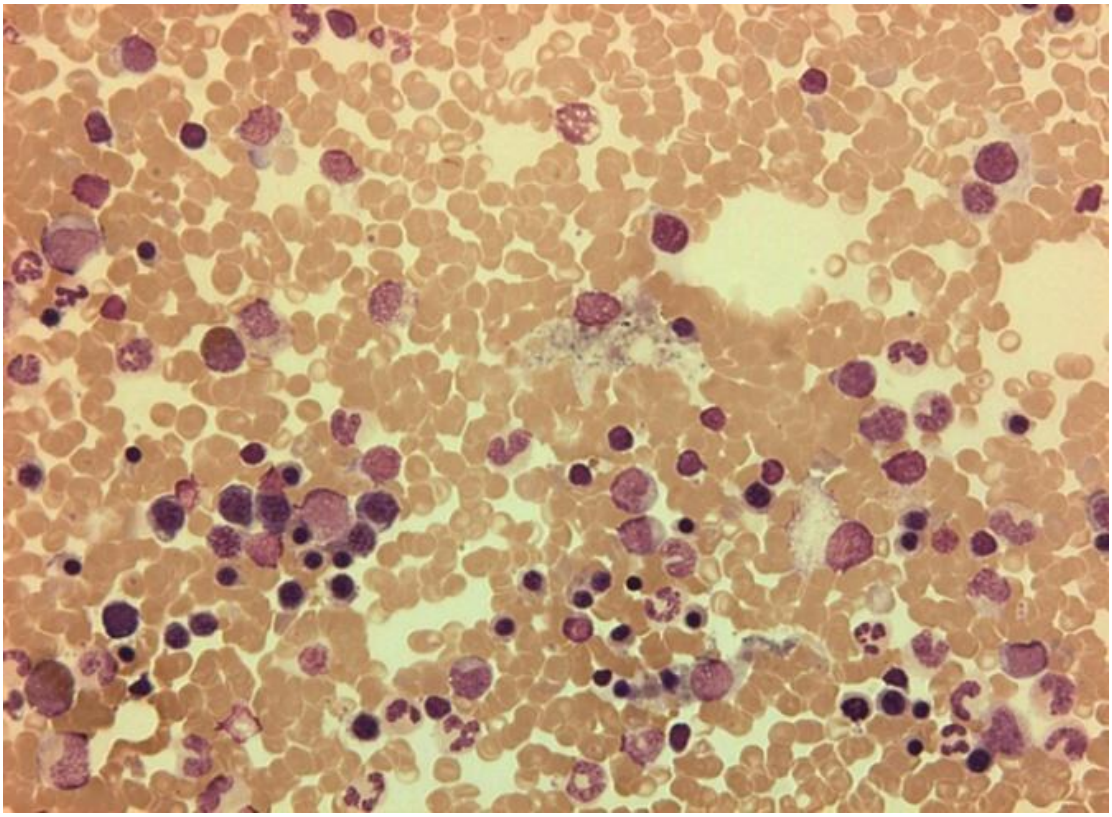


Figure 2.2-2 Bone marrow aspirate from individual II-1

In 1973 (at the age of 62), the mother of the index case (designated I-2 in ***Figure 2.2-1***) presented to her physician and was diagnosed with "mild congestive cardiac failure." At this time her Hb had been 21.4 g/dl. In 1976 she complained of attacks of dizziness and the Hb was 22.8 g/dl with a haematocrit of 61.3%. Other haematological indices are shown in ***Table 2.2-1***. The attacks of dizziness persisted although were improved following regular venesection. She reportedly died from "heart failure" in the 1980s and there was no record of any echocardiographic investigations having been performed.

PA O2 saturation	Aortic O2 saturation	PA pressure	Aortic pressure	PCWP	V wave pressure	Coronary arteries
77%	98%	78/55	126/78	21	25	Normal

Table 2.2-2 Right heart catheter study in II-1

The brother of the index case (individual II-3) was also known to have erythrocytosis with an initial red cell mass of 56.6 ml/kg. He was 32 years old in 1976 and was being treated with regular venesection. White cell and platelet counts were normal as were tests of oxygenation and haemoglobin function (***Table 2.2-1***). He reported intermittent headaches associated with fortification spectra, light sensitivity and nausea which were alleviated following venesection which was performed intermittently lifelong to maintain a haematocrit <50%. He subsequently developed coeliac disease (anti-endomysial antibodies and scalloped duodenum detected following presentation with dermatitis herpetiformis); paroxysmal atrial fibrillation (treated with warfarin and amiodarone); hypothyroidism; and ischaemic heart disease, with an anterior myocardial infarction associated with left anterior descending artery stenosis which was stented in 1998. An echocardiogram at this time revealed an enlarged left ventricle with mild mitral and tricuspid regurgitation and an estimated right atrial/right ventricular pressure difference of 40 mmHg, indicating at least mild to moderate pulmonary arterial hypertension (the jugular venous pressure was not documented at this time). Subsequent echocardiography

performed in late 2006 to assess increasing shortness of breath demonstrated impairment of left ventricular function and increased pulmonary arterial resistance with pulmonary hypertension classed as moderate. His exercise tolerance declined with increasing shortness of breath and he died in 2007 at the age of 63 while awaiting further investigation for pulmonary hypertension. At this time he had received venesection to maintain haematocrit < 50% and his last EPO level had been 1980 mIU/ml. His karyotype was normal on cytogenetic examination.

Erythrocytosis in the two children of II-3 (individuals III-7 and III-8) was identified in childhood when they were screened for the condition. They both exhibited elevated haematocrit with elevated erythropoietin levels identified subsequently (see ***Table 2.2-1***). Individual III-7, born in 1974, has been generally well and has had three children (designated IV-6, IV-7 and IV-8, born in 1994, 1999 and 2003 respectively) and 2 miscarriages. Child IV-6, a boy, complained of headaches at the age of 5 and was noted to have erythrocytosis with Hb of 16.1 g/dl at that time (mean Hb in children between 2 and 6 years is 12.5 g/dl). He did not provide a DNA sample, although his siblings (individuals IV-7 and IV-8 who have normal Hb and erythropoietin levels) did.

Individual III-1, the son of the index case, was born in 1981 and was noted to have erythrocytosis (Hb 20.2 g/dl) with an inappropriately non-suppressed erythropoietin level of 13.2 mIU/mL.

2.3 Genetic Investigation

The *VHL* and *PHD2* genes were first sequenced in two family members prior to the start of this study and no mutations had been found in the coding or flanking intronic sequences of either gene, implying that the cause of the erythrocytosis was not due to a coding mutation in either of these genes. Blood for DNA extraction was collected from additional consenting family members and DNA was extracted. In view of the well-defined pathway regulating erythropoietin production and the limited number of known proteins participating in it (see ***Table 2.3-1***), I performed a genome-wide linkage study in order to identify whether a common haplotype was inherited by all affected individuals at the locus of each of these candidate genes.

Candidate Genes	
PHD2*	CUL2
VHL*	ELONGIN C
PHD1	ELONGIN B
PHD3	EPO
FIH	EPOR
HIF1α	JAK2
HIF2α	STAT2
HIF3α	

Table 2.3-1 Candidate genes for erythrocytosis
EPOR (Erythropoietin Receptor), JAK2 and STAT2 mutations would be predicted to cause erythrocytosis with suppressed EPO levels. *Sequenced with no coding sequence mutation detected prior to linkage study.

With the available DNA, maximum possible LOD scores for markers with different recombination fractions with the condition in this pedigree was estimated using FastSLINK (Weeks, Ott et al. 1990). This yielded a maximum predicted LOD of 1.81 for a marker with a recombination fraction (Θ) of 0 (i.e. at a genetic distance 0 cM from the diseases-causing locus).

Although this would not be sufficient to prove linkage from an unbiased standpoint, it was hoped that a large proportion of the candidate genes would be excluded using this approach.

Genotyping was performed by Dr C Mein (Queen Mary and Westfield College, University of London) using the Illumina Linkage IV panel which consists of bi-allelic markers distributed across the genome and separated by an average of 0.65cM (coverage is summarized in ***Figure 2.3-1***). Mean heterozygosity (in the Caucasian population) is 43%.

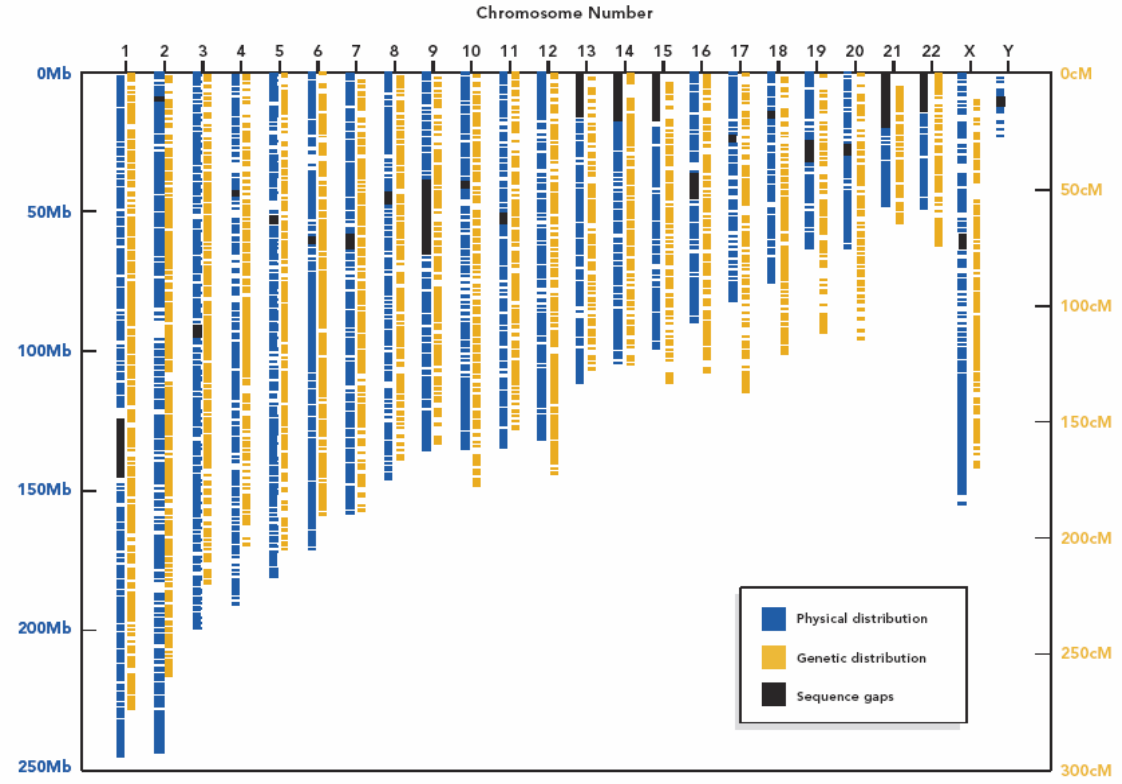


Figure 2.3-1 Genome coverage of Illumina Linkage panel

6008 SNPs were assayed in 9 individuals, yielding 54,072 genotypes which were called using Illumina software. In 16 SNPs, genotyping failed in all individuals. For the remaining 5992 SNPs, genotyping failed in one or more individuals on a further 569 occasions, distributed as shown in ***Table 2.3-2***, yielding a total assay failure rate of 1.3%.

Number of individuals	Number of failed SNPs
1	135
2	17
3	10
4	5
5	7
6	50
7	1
8	1
9 (all)	16

Table 2.3-2 Distribution of SNP genotyping failures.

Genotypes were analysed using PEDCHECK (O'Connell and Weeks 1998)and this indicated that there were 33 Mendelian errors (in 33 different SNPs). This equated to approximately 0.06% genotypes and indicated firstly that the SNP genotyping was very accurate, and secondly that ascertainment of the family tree and sample labelling/processing was likely to be without error. SNPs in which there were Mendelian errors were excluded from all further analyses as these errors were most likely explained by genotyping inaccuracy.

The remaining 53,062 genotypes were analysed using GENEHUNTER (Markianos, Daly et al. 2001) using the EasyLINKAGE interface (Hoffmann and Lindner 2005). The genotypes were divided into sets of 100 markers and parametric linkage analysis was performed in all individuals on each set sequentially, assuming a dominant model of inheritance. The analysis was repeated in sets of 71 and 59 markers and using MERLIN (Abecasis, Cherny et al. 2002). In all analyses, the results were essentially similar.

2.3.1 Genome-wide SNP linkage and haplotype analysis

5522 SNPs (or 92.7% of those analysed) lay in regions of the genome at which the LOD score was less than -2.0 (robustly excluding linkage with erythrocytosis by conventional criteria), implying a >100:1 likelihood against linkage with these loci (at Θ = 0). As predicted, the maximum LOD score was 1.81: 121 SNPs (or 2% of all genotyped) lay in 4 regions of the genome at which the LOD score was >1.5, and a further 54 (0.9%) SNPs lay in regions at which the LOD was between 0 and 1.5. The LOD scores across the entire genome, generated using both GENEHUNTER and MERLIN, are summarized in ***Figure 2.3-2***. The LOD scores at the specific genomic locations of each of the candidate genes are shown in ***Table 2.3-3***. This indicated that, while most of the candidate genes were robustly excluded by conventional criteria (LOD < -2), three genes (*HIF2α*, *VHL* and *Elongin B*) lay in regions which were not formally excluded.

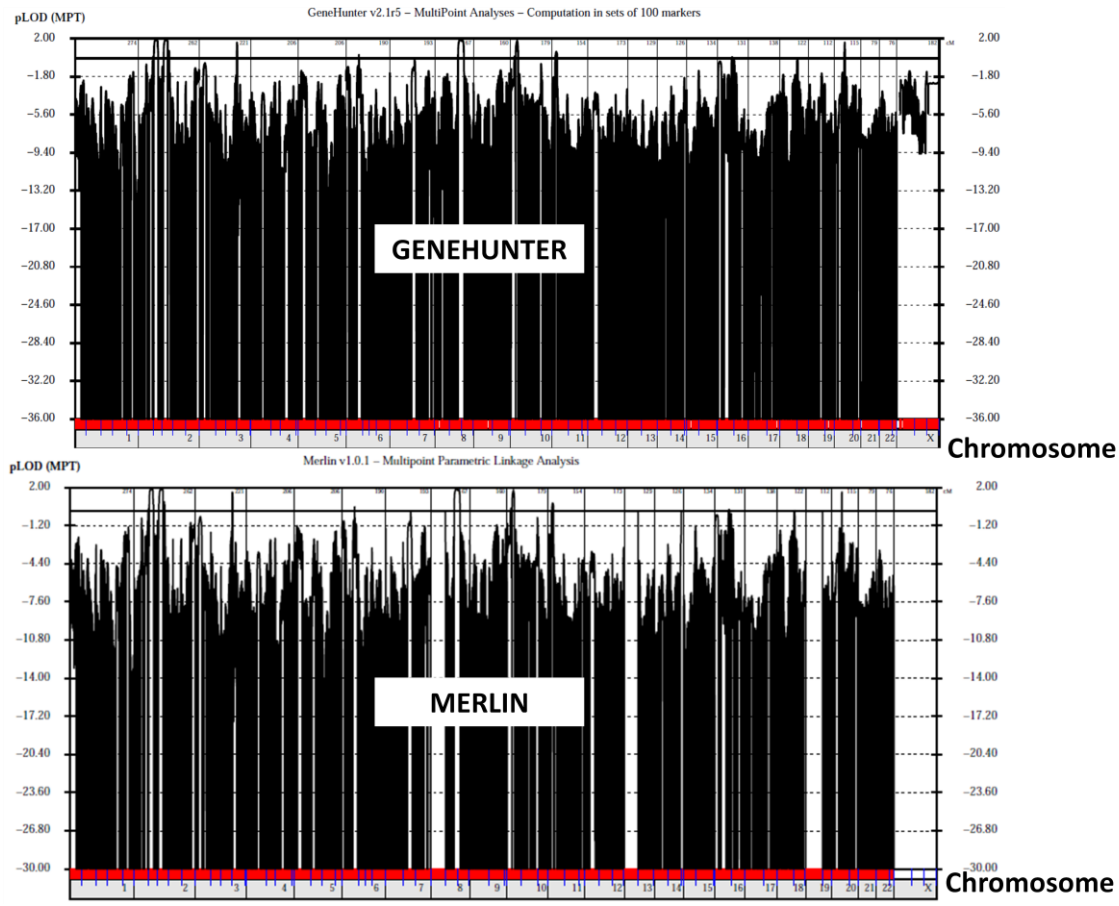


Figure 2.3-2 Genome wide LOD plots with GENEHUNTER and MERLIN

I next performed haplotype analysis using the haplotypes generated by GENEHUNTER and visualised using HAPLOPAINTER (Thiele and Nurnberg 2005). This demonstrated that, of the three genes, only HIF2α lay in a region in which the same haplotype was inherited by all affected members of the pedigree, and hence the LOD score was maximal (***Figure 2.3-3***).

Gene	LOD	Gene	LOD
PHD2	-2.50	CUL2	-5.64
VHL*	-0.91	ELONGIN C	-10.71
PHD1	-5.10	ELONGIN B*	-0.41
PHD3	-5.81	EPO	-4.84
FIH	-6.70	EPOR	-10.85
HIF1α	-9.75	JAK2	-6.69
HIF2α**	**1.81**	STAT2	-11.51
HIF3α	-11.7		

Table 2.3-3 Parametric LOD score at each candidate gene
*Not formally excluded by conventional criteria
**Only candidate gene at locus which segregates perfectly with disease

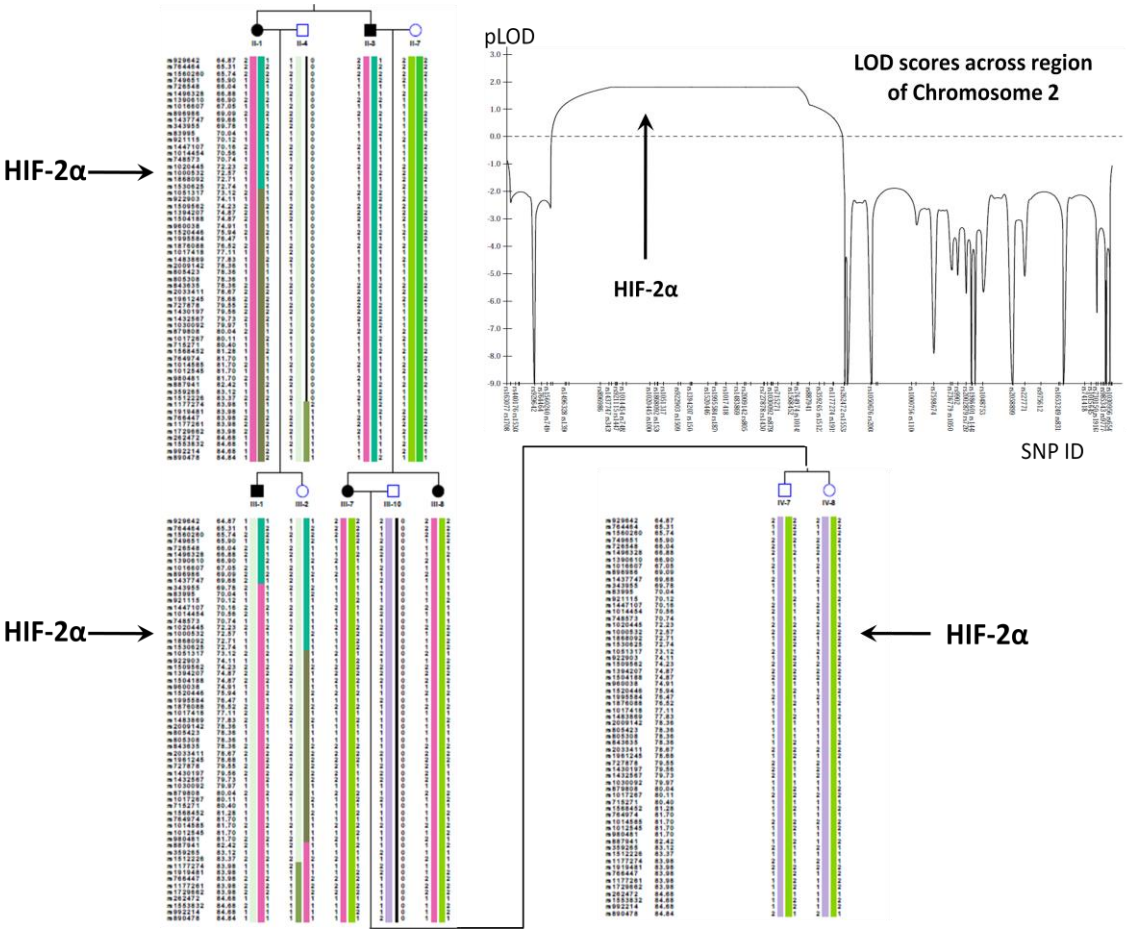


Figure 2.3-3 Haplotype analysis

A haplotype (pink) which spans the HIF2α locus is inherited by all (and only) the affected individuals in the family (filled symbols). The LOD plot was generated using GENEHUNTER scoring unaffected individuals as 'unknown'

I concluded that the mutation responsible for the disease was likely to lie within the HIF2α gene.

2.3.2 HIF2α *sequencing*

The *HIF2α* gene consists of 5160 base-pairs (bp) divided between 16 exons in a genomic region of approximately 42 kilobase-pairs (kbp). However, since the phenotype of the disease suggested *increased* HIF activity, I hypothesised that to produce such a gain-of-function, the mutation responsible, given that it lay within the *HIF2α* gene, would be more likely to disrupt the function of a regulatory domain than to enhance the function of an activation domain. This is because a random change in a complex system is more likely to impair its function than to enhance it. An analogy is shooting a bullet into a motor car. This will either result in no effect

on the function of the car or disruption of its function: either being stopped when it should be going or going when it should be stopped. If it is the latter, then the bullet would be intuitively more likely to have damaged the brakes rather than to have changed the engine in such a way as to increase its performance. In addition, experimental manipulation of the HIF pathway in cell-based models has been performed employing mutations of the critical proline and asparigine residues (to alanines) in order to generate HIF proteins which are resistant to oxygen-dependent degradation (Kondo, Klco et al. 2002). For these reasons, it seemed likely that the mutation responsible would lie at or near proline residues 405 or 531 or else the asparagine residue 847.

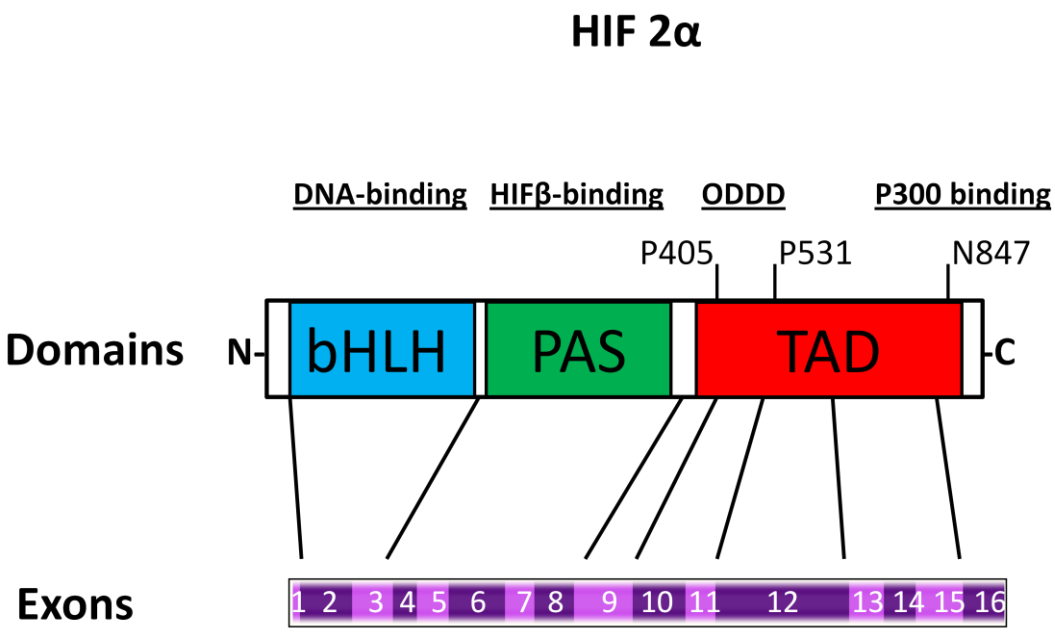


Figure 2.3-4 Exons of HIF-2α mapped to functional domains of the protein
ODDD, oxygen-dependent degradation domain; bHLH, basic helix-loop-helix; PAS Per-Arnt-Sim; TAD, Terminal activation domain

I therefore mapped the exons of the *HIF2α* gene onto the various functional domains of the HIF2α corresponding to the oxygen dependent degradation domain (ODDD) and P300 binding domain (***Figure 2.3-4***) and amplified and sequenced exons 9, 12 and 16 of the gene in 2 affected members of the family (II-1 and III-7) and one control person. This demonstrated

heterozygosity for a Guanine to Adenine (G to A) single nucleotide substitution at base 2097 from the transcription start site in both affected (but not the unaffected) individuals (***Figure 2.3-5***). This substitution was not represented in any online database of single nucleotide polymorphisms.

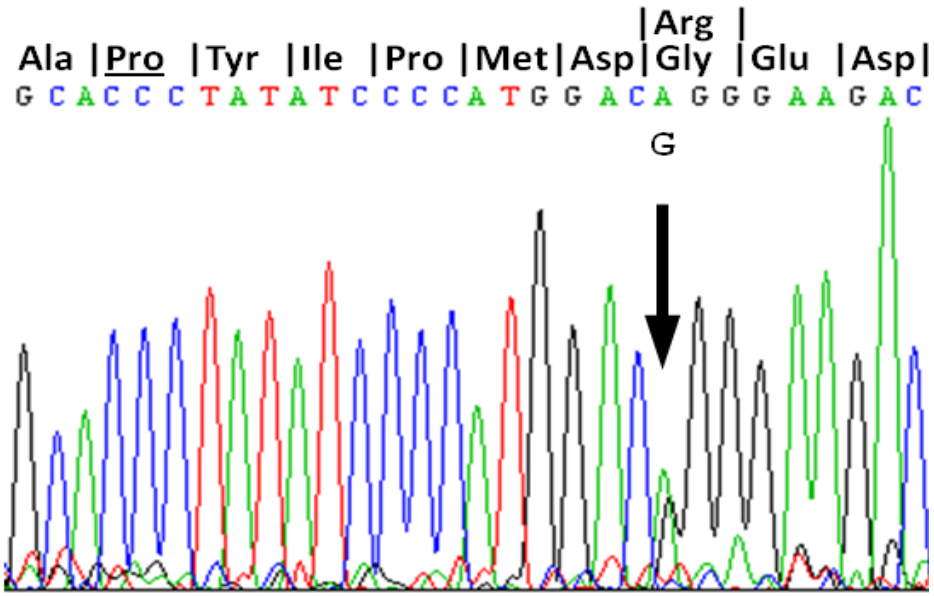


Figure 2.3-5 Sequence trace
Resequencing affected individuals II-1 and III-7 revealed heterozygosity for a G→A substitution at base 2097 in exon 12 of *HIF-2α* (arrow). This is not a previously reported SNP and predicts a Glycine (Gly) to Arginine (Arg) substitution at residue 537 of the HIF2α protein, 6 amino acids from the functionally important hydroxyl acceptor Proline (Pro – underlined).

2.3.2.1 HIF2α *G2097A may be pathogenic*

HIF2α G2097A is predicted to be a non-synonymous change: it predicts a change from the small and hydrophobic Glycine (G - wild type) to the large and hydrophilic Arginine (R). Notably, this residue is highly conserved across species (***Figure 2.3-6***) and lies just 6 residues from Proline 531, which is one of the two hydroxyl-acceptor prolines of the ODDD. It therefore seemed possible that this substitution could impair the hydroxylation of HIF2α or else reduce the ability of the VHL E3 ubiquitin ligase to recognise and/or ubiquitinate the hydroxylated protein, reducing the rate of its degradation.

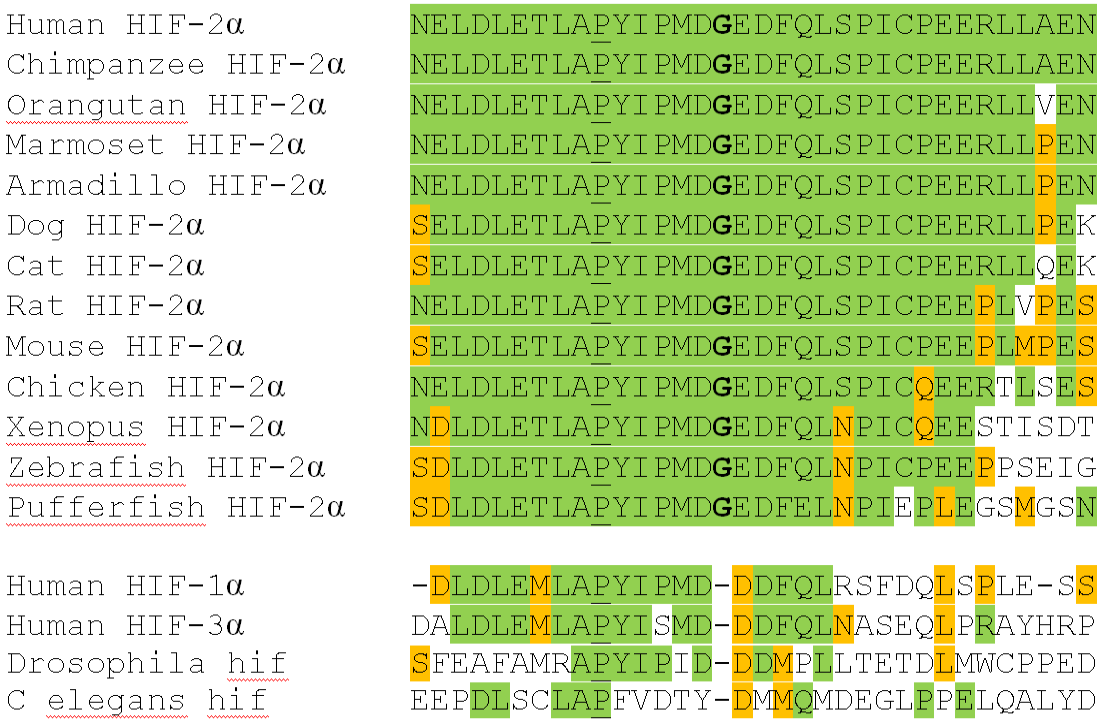


Figure 2.3-6 Conservation of HIF2α537Gly

537Gly (bold) is highly conserved across many species in HIF2α but is not present in HIF1α. Hydroxyl-acceptor Proline (P) is underlined

2.3.2.2 HIF2α *G2097A is a rare mutation which segregates with disease*

The substitution G2097A does not produce or abolish any known restriction sites, so in order to determine whether it was present in other members of the pedigree or in healthy controls I used an amplification refractory mutagenesis system PCR (ARMS-PCR) reaction. This technique was originally described in 2001 (Ye, Dhillon et al. 2001) and uses 2 primer pairs, each selective for one allele, to generate products of different sizes. The primers compete and the products can be separated by gel electrophoresis. This confirmed firstly that the mutation did cosegregate with the disease (***Figure 2.3-7***) and secondly that the allele was not present in 88 members of the UK 1958 birth cohort (***Figure 2.3-8***). Together these findings supported the hypothesis that the G2097A mutation was responsible for erythrocytosis and pulmonary arterial hypertension in the family.

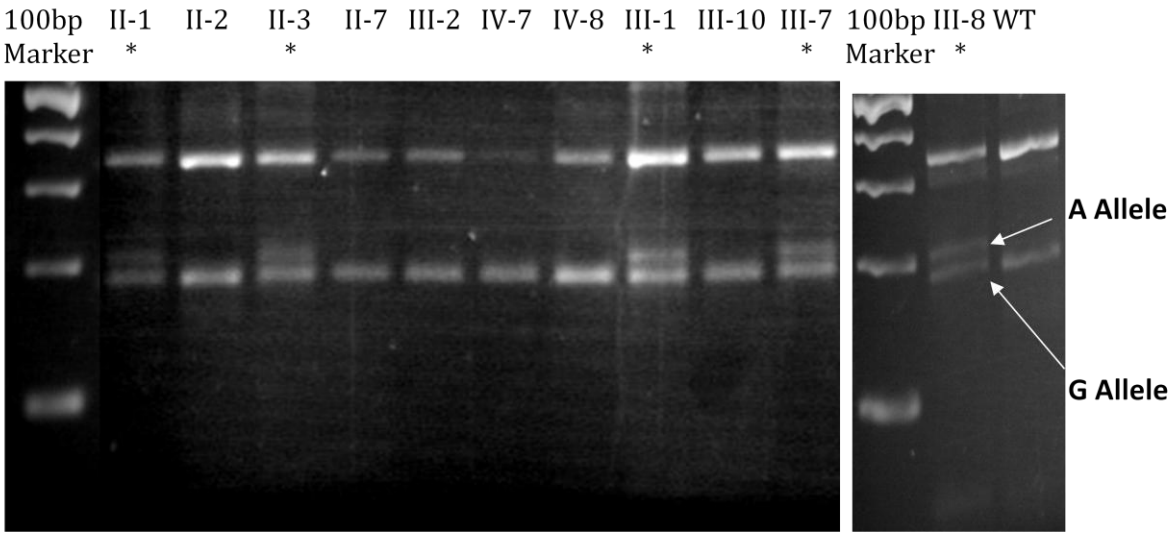


Figure 2.3-7 ARMS-PCR assay
Demonstrating that the G2097A heterozygosity cosegregates with the disease within the family.

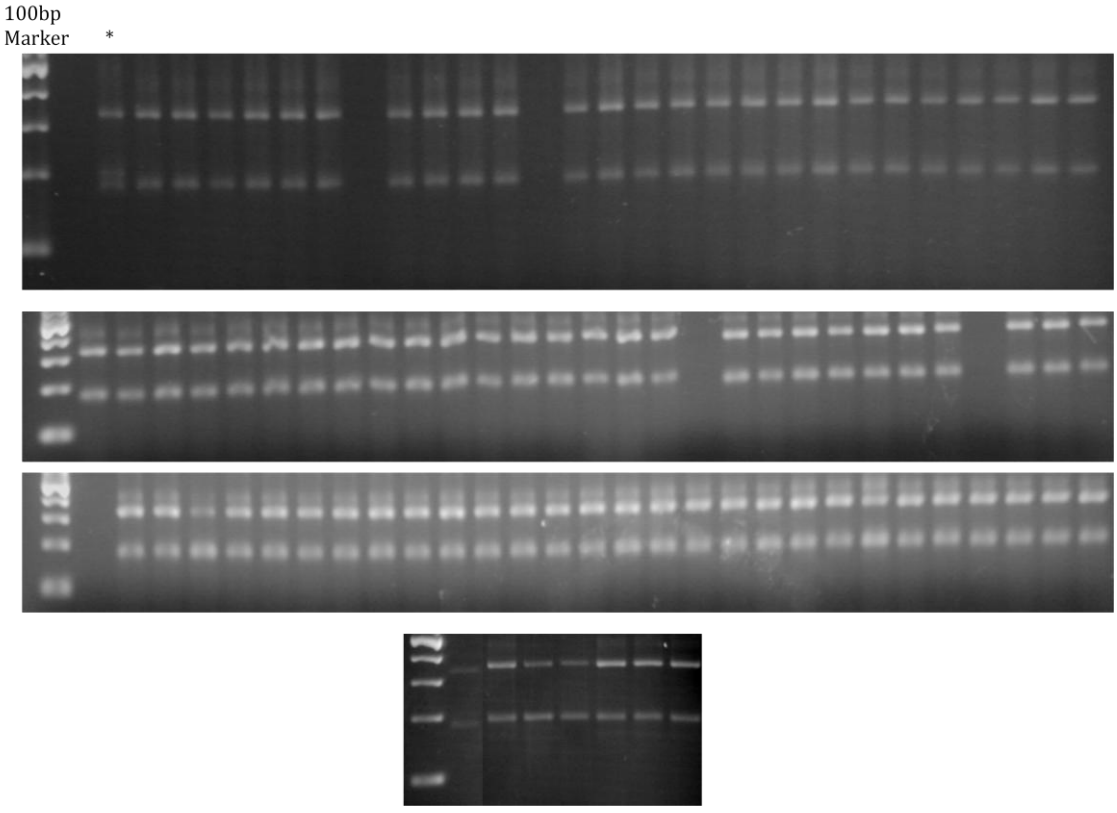


Figure 2.3-8 ARMS-PCR in 88 UK controls
Demonstrating lack of heterozygosity for G2097A in HIF2α among 88 members of the UK 1958 birth cohort (selected at random). Lane 1 (asterisk in first panel) is individual II-1, used as a positive control.

2.4 HIF2α Arg537 functional studies

HIF2α 2097A codes for a protein identical to the wild type HIF2α with the exception of the substitution of an Arginine for the Glycine at position residue 537, 6 amino acids away from the critical hydroxyl acceptor proline. In order to test whether this substitution is sufficient to alter the function of the protein, I generated the mutant alleles *in vitro* using site-directed mutagenesis and used an HRE-luciferase reporter system in order to determine whether the mutant protein exhibited altered function compared with the wild type.

2.4.1 HIF2α Arg537 has increased transcriptional activity

At this time, a report of a different small family was published in which erythrocytosis and raised EPO levels cosegregated with a mutation which resulted in a Glycine to Tryptophan change at position 537 of HIF2α (i.e. a different mutation at the same codon as the mutation which I had identified) (Percy, Furlow et al. 2008). Pulmonary arterial hypertension was not reported in this family.

I therefore generated constructs containing *HIF2α* with either the wild type sequence, or Arginine (Arg537) or Tryptophan (Trp537) instead of Glycine at position 537. I co-transfected these constructs, alongside an HRE-Firefly luciferase reporter construct (in which expression of a firefly luciferase gene is dependent on HIF) and a constitutively expressed *Renilla* luciferase construct, into cultured human hepatoma cells (HepG2) which produce EPO in response to hypoxia (i.e. cells in which oxygen dependent expression of EPO and HIF can be studied (Goldberg, Glass et al. 1987)). The ratio of firefly to *Renilla* luciferase activity (which can be separately measured) in lysates made from these cells is therefore dependent on HIF activity within the cells. This experiment is represented in ***Figure 2.4-1*** and was performed with increasing doses (25, 50 and 100ng) of plasmid DNA, each time in triplicate. As a positive control, I co-transfected cells with the luciferase constructs and incubated cells co-transfected

with the different *HIF2α* alleles with desferrioxamine (DFO), which chelates iron, inhibiting prolyl- and asparaginyl-hydroxylation and therefore stabilising HIF. As a negative control I co-transfected cells with the luciferase constructs and equimolar amounts of the empty pcDNA3.1 vector.

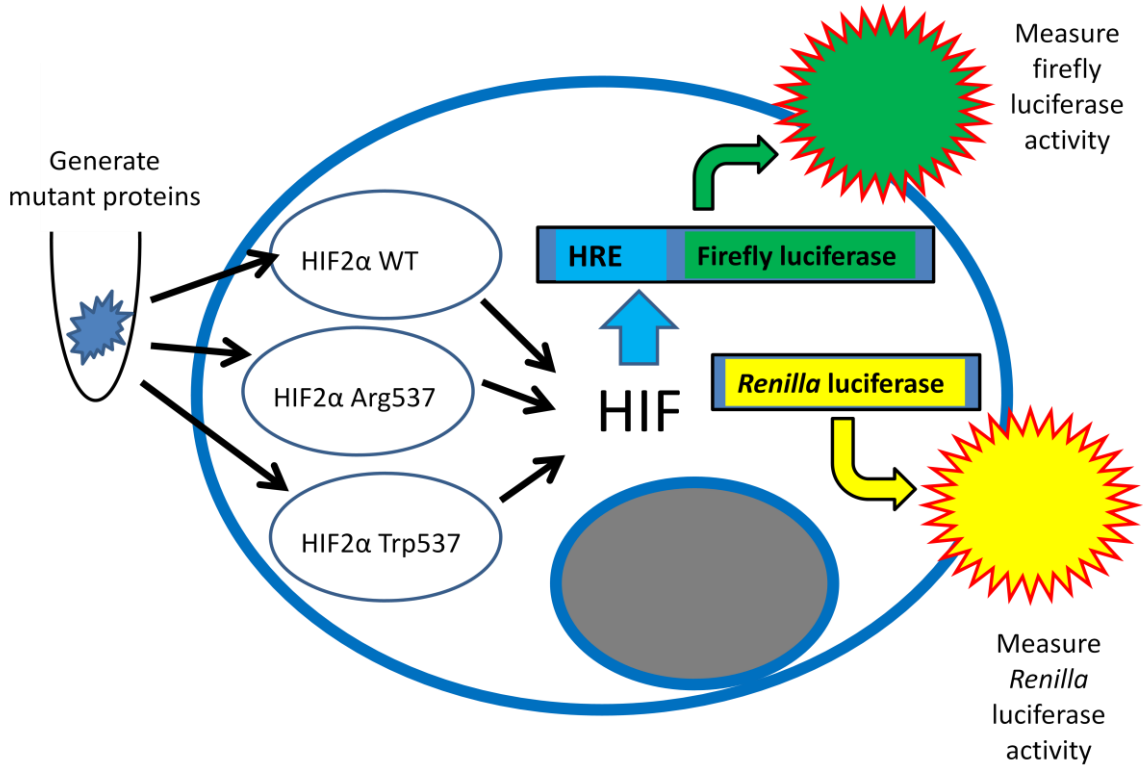


Figure 2.4-1 HIF activity reporter assay
Wild type or mutant HIF2α in pcDNA3.1 was transfected into hepatoma (HepG2) cells alongside constructs containing an HRE-linked (i.e. HIF responsive) firefly luciferase gene and a constitutively expressed *Renilla* luciferase gene. The ratio of firefly:*Renilla* luciferase activity is dependent on the amount of HIF activity within the cell.

In all experiments, the Firefly:*Renilla* luciferase ratio was greater in cells co-transfected with the same quantity of the Arg537 mutant HIF2α construct compared with the wild type. Using a construct containing HIF2α Trp537 instead of Arg537 resulted in intermediate luciferase expression (***Figure 2.4-2***).

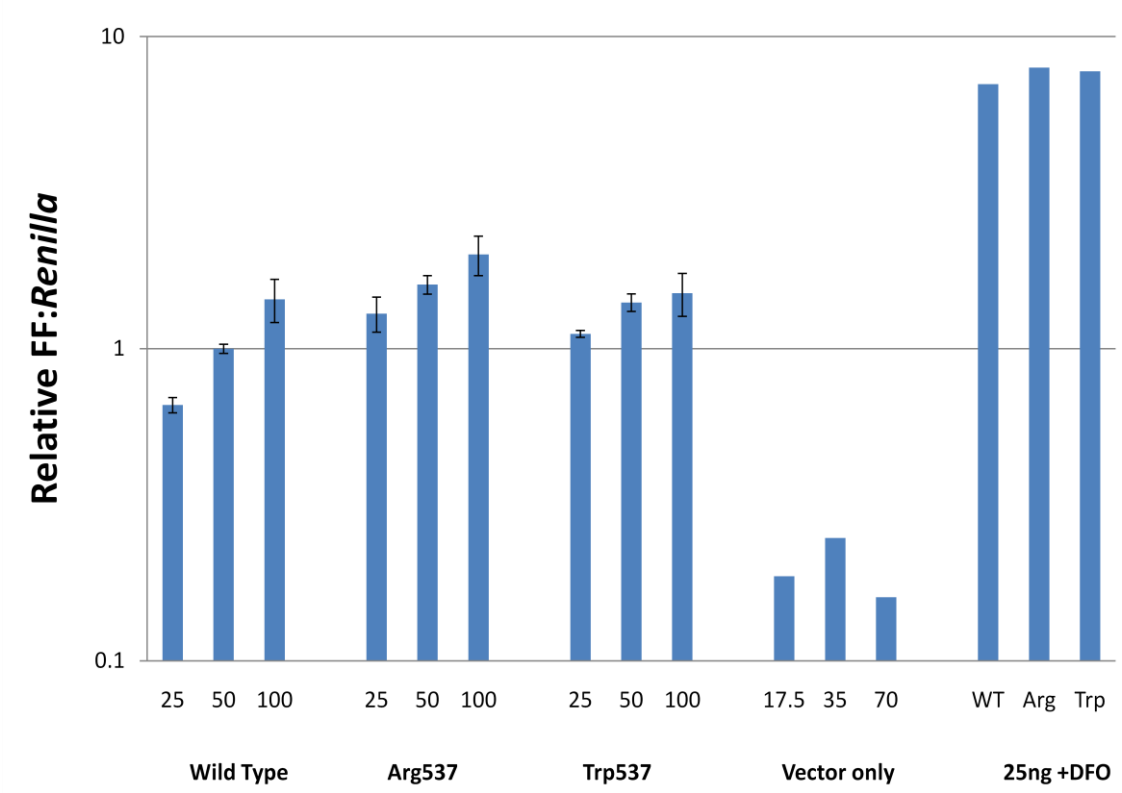


Figure 2.4-2 HIF2α activity

Firefly (FF):Renilla luciferase activity in HepG2 cells transfected with an HRE-Firefly luciferase construct and a *Renilla* luciferase construct alongside either: plasmid constructs containing HIF-2α wild type, Arg537 or Trp537 mutants at the quantities shown in ng in triplicate; or equimolar amounts of the empty vector; or 25ng of the wild type and mutant constructs and incubated with 100μM desferrioxamine (DFO). Error bars represent standard deviation.

The dose-response ratio can be appreciated by representing the same data as a line graph (***Figure 2.4-3***) and the differences were statistically significant at a dose of 50 ng (one way ANOVA: ***Figure 2.4-4***). Stabilisation of HIF with DFO produced significantly greater firefly:*Renilla* ratios than any of the mutants incubated in the absence of DFO. This effect was independent of genotype and there was no significant difference between the different alleles. I repeated the experiment using Hep3B cells (a different hepatoma cell line in which HIF and EPO are regulated by oxygen) and saw similar results (***Figure 2.4-5***).

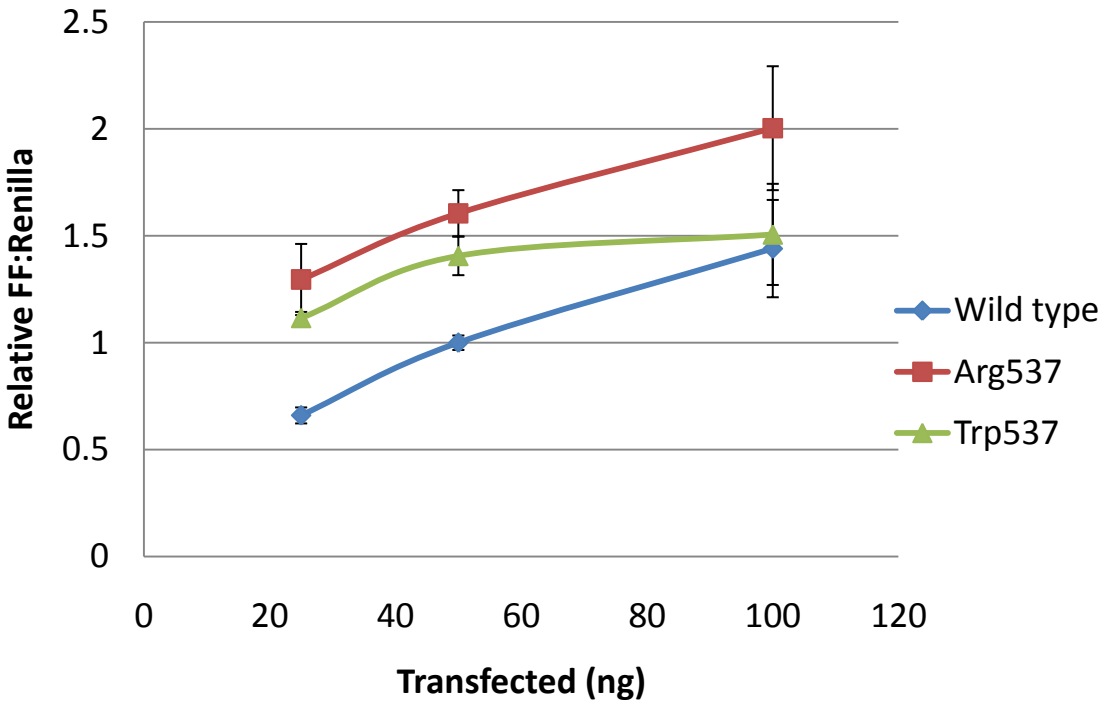


Figure 2.4-3 Dose-dependent reporter activity

Firefly (FF):*Renilla* luciferase activity ratio increased on transfection of increasing amounts of each HIF-2α construct in HepG2 cells. Error bars represent standard deviation.

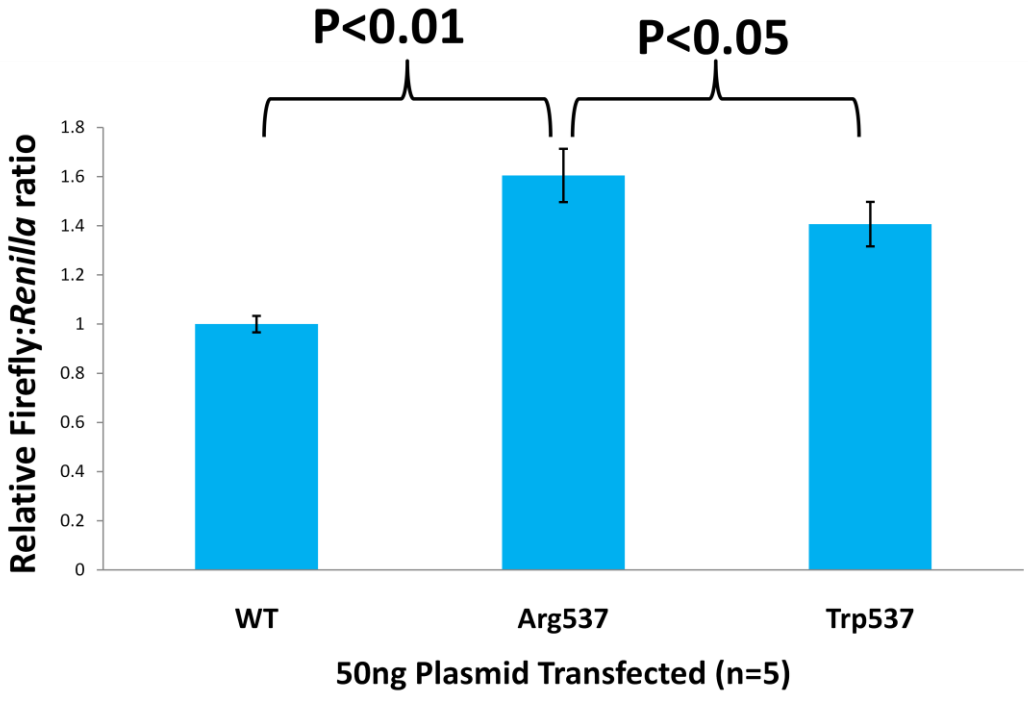


Figure 2.4-4 Activity of each HIF2α allele

50 ng of plasmid containing either mutant or wild type HIF-2α was co-transfected into HepG2 cells alongside HRE-firefly luciferase and *Renilla* luciferase reporter constructs. Transfection with HIF-2α Arg537 resulted in significantly greater Firefly:*Renilla* luciferase ratio than transfection with either HIF-2α wild type or HIF-2α Trp537. Experiment performed 5 times. Error bars represent standard deviation; p = one way repeated measures ANOVA.

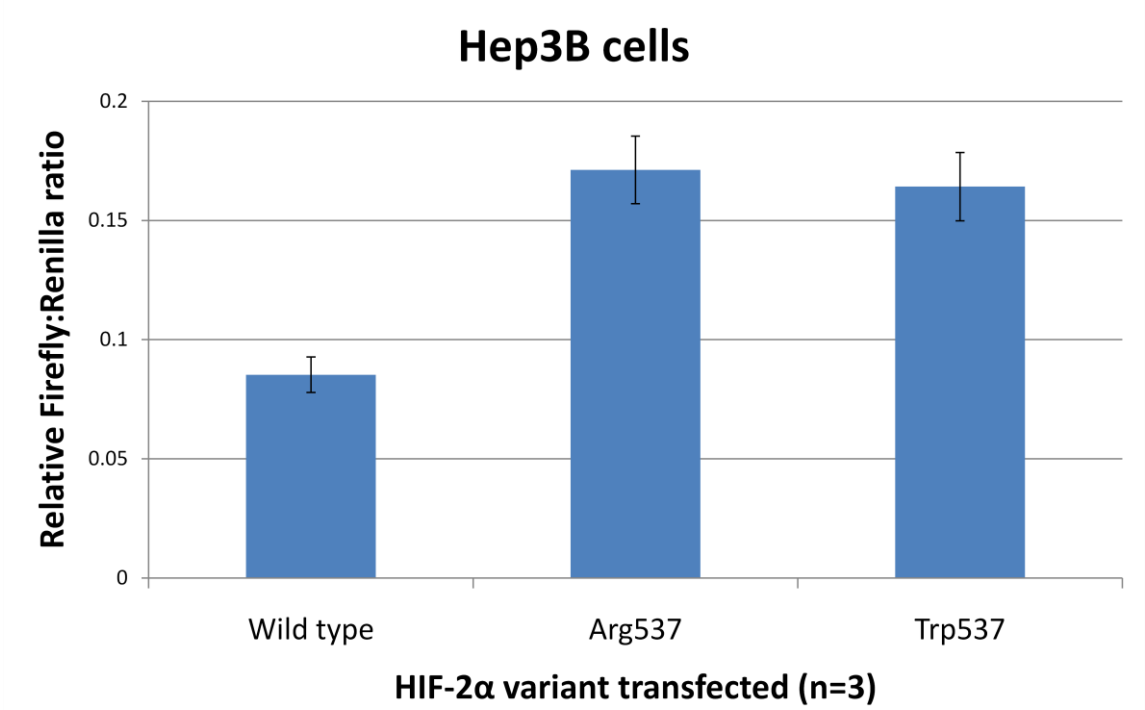


Figure 2.4-5 HRE-firefly:*Renilla *luciferase activity in Hep3B cells

2.4.2 Increased transcriptional activity of HIF2α Arg537 is oxygen-dependent

In order to identify whether the increased reporter gene expression (i.e. increased mutant HIF activity) was dependent on the presence of oxygen, I performed a further experiment in which the transfected HepG2 cells were incubated at various oxygen tensions for 16h immediately prior to their lysis. This demonstrated that, as the oxygen level decreased, the difference between the genotypes became less marked. At 1% oxygen (a level at which prolyl- and asparginyl-hydroxylation of wild type HIF is very inefficient so that HIF is highly active) there was no difference between the activities of the different HIF2α proteins (***Figure 2.4-6***). I concluded that the mutation impaired the oxygen-dependent hydroxylation of HIF2α and was unlikely to have a significant direct effect on DNA- or HIF1β-binding or co-transactivator recruitment.

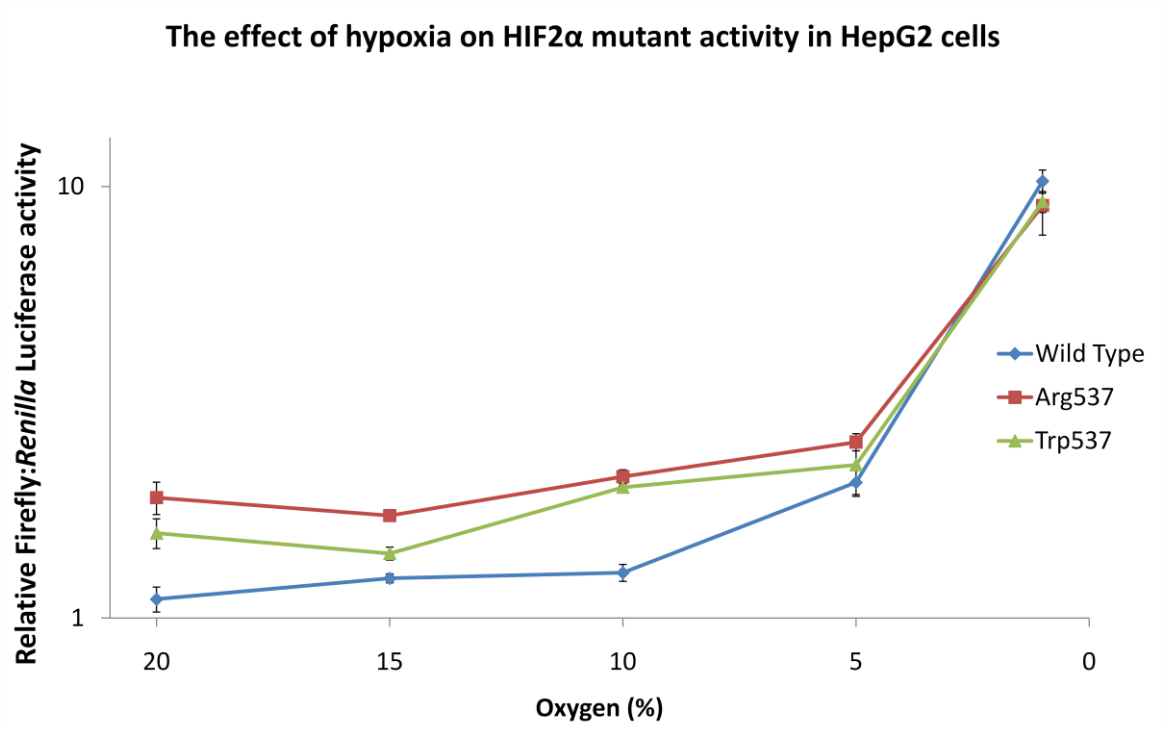


Figure 2.4-6 Effect of hypoxia on HIF2α mutant activity in HepG2 cells
In hypoxia, activity of all constructs was increased and the difference between the level of activation was no longer detectable.

2.5 HIF2α Erythrocytosis and pulmonary physiology

The observation that two older members of the pedigree had pulmonary arterial hypertension (in the absence of pulmonary thromboembolic disease) was potentially significant because evidence already existed which implicates the HIF pathway in the regulation of pulmonary vascular tone (see introduction to this chapter). In order to identify whether HIF2αArg537 causes elevated pulmonary vascular tone, I studied two unrelated individuals with the mutation. The first was individual III-1 who, at the age of 26, was polycythaemic with a haemoglobin level of 20 g/dl. He had never been venesected and was generally in good health with the exception of occasional headaches.

The second individual, designated DC, was a 44-year-old man of mixed Indian and Kenyan ancestry who had been diagnosed with polycythaemia aged 21 while living in Uganda. He had undergone regular venesection to maintain a haematocrit below 50% since then. Sequencing of

exon 12 of HIF2α performed by others had demonstrated heterozygosity for G2097A, predicting the same Glycine to Arginine change at residue 537 of the protein (Percy, Beer et al. 2008).

The experimental protocol was the same as that described previously (Smith, Brooks et al. 2006). In brief, end tidal partial pressure of oxygen and carbon dioxide (PET_{O2} and PET_{CO2}) were measured by occluding the nose and allowing breathing to occur through a mouthpiece connected to a circuit. Continuous monitoring of exhaled respiratory gases was performed using a mass spectrometer connected to a computer which also controlled the delivery of nitrogen, oxygen and carbon dioxide to the subject, allowing dynamic control of the PET_{O2} and PET_{CO2}. PET_{CO2} was clamped at the subject's own baseline level throughout the experiment. Each participant was first subjected to 5 minutes of euoxia (PET_{O2} of 100 mmHg) followed by 10 minutes of mild hypoxia (PET_{O2} of 70 mmHg, approximating to the level of hypoxia experienced during commercial airline travel). This was followed by 5 minutes of euoxia before a 20 minute break breathing room air. Next the protocol was repeated but with the hypoxic stimulus being moderate instead of mild i.e. a PET_{O2} of 50 mmHg, which is approximately equivalent to the hypoxia experienced during exposure to an altitude of 3500m above sea level. The protocol was performed with the subject in the left lateral position and a Doppler ultrasound machine was used to measure the maximum systolic tricuspid regurgitant jet velocity, V. The maximum systolic pressure gradient (ΔP_{max}) across the valve (i.e. the difference between the pulmonary arterial pressure and the right atrial pressure) was calculated using Bernoulli's equation:

$$\Delta P_{max} = 4V^2$$

Since right atrial pressure is not affected by hypoxia (Groves, Reeves et al. 1987), changes in ΔP_{max} can be attributed to changes in pulmonary arterial resistance.

This experiment demonstrated firstly that resting ΔP_{max} was elevated in both individuals with HIF2α erythrocytosis at 30 (III-1) and 29 mmHg (DC). Since the cut-off for pulmonary arterial hypertension is 25 mmHg (and ΔP_{max} is likely to underestimate the PA pressure since the right atrial pressure is greater than 0 mmHg) both these individuals have pulmonary arterial hypertension at rest. In addition, mild and moderate hypoxic exposure provoked an exaggerated rise in ΔP_{max} in both individuals compared with healthy controls (***Figure 2.5-1***). I concluded that activation of HIF2α alone (i.e. in the absence of activation of HIF1α) is sufficient to produce exaggerated pulmonary arterial tone in humans.

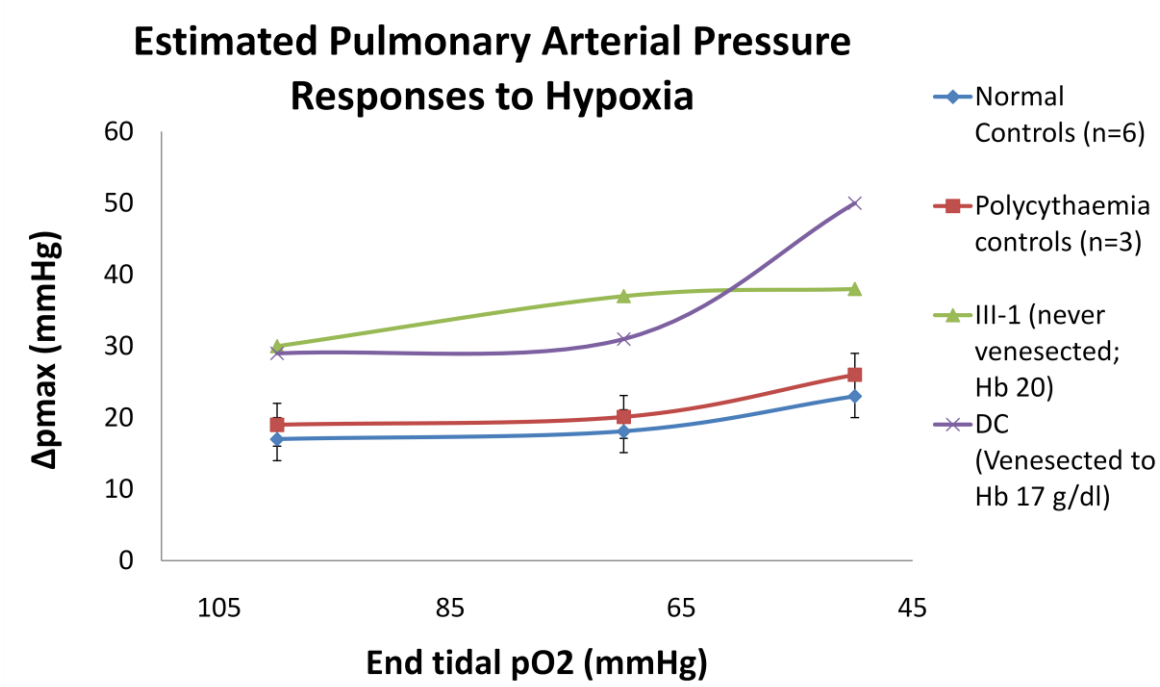


Figure 2.5-1 Isocapnic hypoxic exposure in HIF2α Erythrocytosis
Hypoxic pulmonary vasoconstriction responses were exaggerated in individuals with HIF2α Erythrocytosis compared with controls.

2.6 Summary

This investigation resulted in the identification of a novel mutation in the *HIF2α* gene which cosegregated with erythrocytosis and elevated pulmonary arterial pressure. Experiments

performed using recombinant mutant and wild-type protein demonstrated that the substitution of Arginine (or Tryptophan) in the place of Glycine at residue 537 of HIF2α results in a protein in which transcriptional activity in normoxia is enhanced. This is very likely to be a consequence of impaired oxygen-dependent degradation of the protein since the difference is not seen when HIF prolyl hydroxylation is inhibited (by either iron chelation or hypoxia) and is almost certain to cause the disease.

2.7 Discussion

This investigation led to the identification of an activating mutation in *HIF2α* which was responsible for autosomal dominant erythrocytosis and pulmonary arterial hypertension in a UK family. Consistent with this, I showed *in vitro* that the mutation resulted in increased HIF2α activity in the presence of oxygen.

The role of HIF2α in regulating erythropoietin production is supported by several previous observations. Firstly, selective knockdown of HIF2α but not HIF1α using RNA interference in hepatoma cell lines caused loss of hypoxia-induced upregulation of the erythropoietin gene. This contrasted with the dependency of hypoxia-induced expression of most other HIF target genes tested on HIF1α (Warnecke, Zaborowska et al. 2004). Secondly, conditional inactivation of *hif2α* but not *hif1α* in the livers of mice deficient in *vhl* suppressed hepatic expression of erythropoietin and the development of polycythaemia (Rankin, Biju et al. 2007). Thirdly, expression of *hif2α* but not *hif1α* in ischaemic rat kidneys was localised to the fibroblasts which produce erythropoietin (Maxwell, Osmond et al. 1993; Rosenberger, Mandriota et al. 2002). In addition, mice homozygous for the *vhl* R200W mutation which causes Chuvash polycythaemia in humans have been shown to have increased haematopoiesis associated with increased *hif2α* protein levels (Hickey, Lam et al. 2007). Together these previous findings are all compatible with a crucial role for HIF2α in the control of erythropoietin production in humans.

While the study presented here was being done, a separate group identified unrelated individuals with mutations in the same and nearby codons of *HIF2α* which are also associated with erythrocytosis (Percy, Beer et al. 2008; Percy, Furlow et al. 2008). This proves that activation of HIF2α in humans is sufficient to cause erythrocytosis. Although in mice inducible silencing of *hif2α* results in normocytic anaemia (Yamashita, Ohneda et al. 2008), proof of a non-redundant role for HIF2α in erythropoiesis in humans would require either the identification of a loss-of-function *HIF2α* allele which is associated with a lower than normal

haematocrit, or alternatively observing anaemia in humans following specific targeting of HIF2α with a pharmacological agent.

A genome-wide association study of haemoglobin levels in 16,000 individuals of European or Indian ancestry identified a non-synonymous SNP located in *TMPRSS6* (which codes for an enzyme involved in the regulation of the synthesis of the iron control hormone hepcidin) as the most strongly associated with haemoglobin levels, and other SNPs located nearby and within the *HFE* gene (mutations of which cause hereditary haemochromatosis, or progressive iron loading) were also significantly associated with haemoglobin levels (Chambers, Zhang et al. 2009). Importantly, variants of *HIF2α* or the genes constituting the oxygen sensing pathway were not found to be associated with haemoglobin levels in this study. It appears likely that this simply reflects lack of common functionally significant variation in these genes in the populations examined. This view is supported by recent data obtained from genetic analyses of high altitude populations (see below).

It is well recognised that hypoxia (for instance during altitude exposure) is a risk factor for the development of pulmonary hypertension (Preston 2007). The physiological reason for this phenomenon may be a short term adaptive response: as regions of the lung in which ventilation is impaired (for instance in the context of pneumonia) become hypoxic the increasing vascular resistance through these regions reduces the amount of deoxygenated blood entering the systemic arterial circulation and prevents shunting of deoxygenated blood back to the systemic arterial circulation.

As discussed above, patients with Chuvash polycythaemia exhibit elevated resting pulmonary vascular tone (Bushuev, Miasnikova et al. 2006); exaggerated pulmonary vasoconstriction in response to hypoxic exposure (Smith, Brooks et al. 2006); and mice engineered to carry the same homozygous substitution (R200W in *vhl*) exhibit pathological features of pulmonary arterial hypertension (Hickey, Richardson et al. 2010). This is compatible

with the hypothesis that both the abnormal physiology displayed by patients with Chuvash polycythaemia and the normal human pulmonary vascular response to altitude are mediated by HIF2α. Further insights into the role of the HIF pathway, and HIF2α in particular, have come from studies of individuals and communities exposed to chronic hypoxia which is unavoidable at high altitudes.

2.7.1 Genetic adaptation to altitude

Humans have adapted to a range of different environments since their dispersal from Africa. Behavioural, technological and biological adaptability have all contributed to enable human survival in a vast geographical range which spans areas in which temperature, ultraviolet radiation, climate and food sources differ enormously. However, of these forms of adaptability, only biological (i.e. genetic) adaptation has contributed to the human ability to colonise high-altitude territory, because pre-modern technology could not ameliorate the pervasive hypoxia which exists at high altitude locations as a consequence of reduced atmospheric pressure. Over 140 million people worldwide now live at over 2500m above sea level (Moore, Niermeyer et al. 1998; Penaloza and Arias-Stella 2007), and communities have lived at altitude in the Tibetan plateau and Andean highlands for many generations (estimated at 1100 and 550 respectively).

2.7.1.1 Chronic Mountain Sickness

Exposure to high altitude causes a reduction in the partial pressure of oxygen in the alveoli, which in turn results in reduced saturation of haemoglobin (***Figure 2.7-1***). Since the amount of oxygen dissolved in the plasma is small compared with the amount bound to haemoglobin, most of the oxygen carrying capacity of the circulation is provided by haemoglobin. As the partial pressure of oxygen falls below 100 mmHg this is associated with a steepening of the oxygen-haemoglobin dissociation curve (***Figure 2.7-2***), with any further reduction in oxygen concentration being associated with a more pronounced decline in haemoglobin saturation.

This stimulus is accompanied by a rise in haematocrit and a marked decline in plasma volume, presumably as an adaptive mechanism to increase the oxygen carrying capacity of the blood by maximising the haematocrit as rapidly as possible. It is notable that the plasma volumes of 22.4 and 28.0 ml/kg reported by Howarth et al in 1979 in patients with HIF2α erythrocytosis were lower than any seen in those authors' clinical practice (Howarth, Chanarin et al. 1979). In fact altitude-induced reductions in plasma volumes (of up to 39% and as low as 29 ml/kg) were recorded as long ago as 1964 by LG Pugh who studied the effects of acute altitude exposure on Europeans during a Himalayan expedition (Pugh 1964).

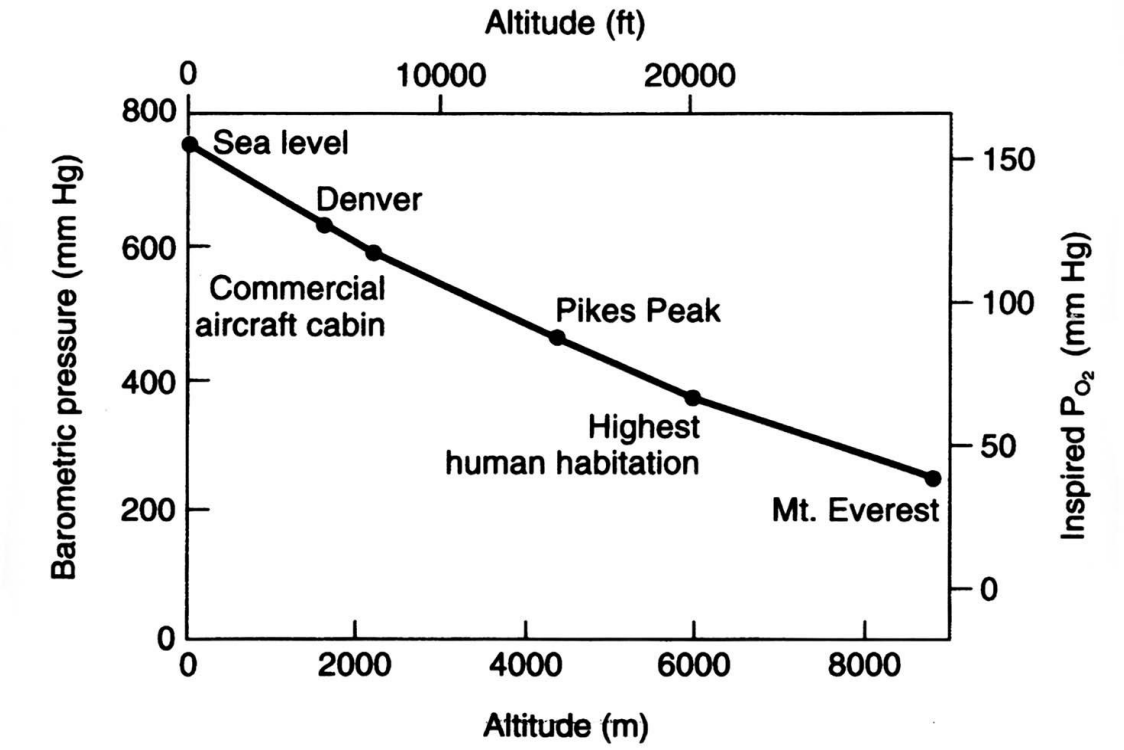


Figure 2.7-1 High altitude is associated with reduced barometric pressure
This results in reduced inspired partial pressure of oxygen

Prolonged exposure to reduced ambient oxygen can result in profound adverse physiological effects known as chronic mountain sickness (CMS). CMS is a syndrome which occurs in long-term residents at high altitude characterised by excessive erythrocytosis

(increase in red blood cell count), hypoxaemia and reversibility on descent. It is well recognised that the progressive CMS leads to pulmonary arterial hypertension with consequent right ventricular enlargement (Moore 2001). This is known as high altitude heart disease (HAHD) and autopsy findings in fatal cases of CMS have demonstrated that the main abnormal pathology involves hypertrophy of the vessels of the pulmonary circulation (Arias-Stella, Kruger et al. 1973). Observations in healthy residents of high altitude have demonstrated that right ventricular hypertrophy, present at birth in all infants, persists in those raised at altitude (Penaloza and Arias-Stella 2007). Importantly, the clinical syndrome of erythrocytosis and pulmonary hypertension is also seen in many patients with chronic hypoxaemia owing to, for example, chronic respiratory or cardiovascular diseases. The similarity of the features of CMS to HIF2α erythrocytosis is compatible with the hypothesis that hypoxic activation of HIF2α underlies the pathophysiology of both diseases.

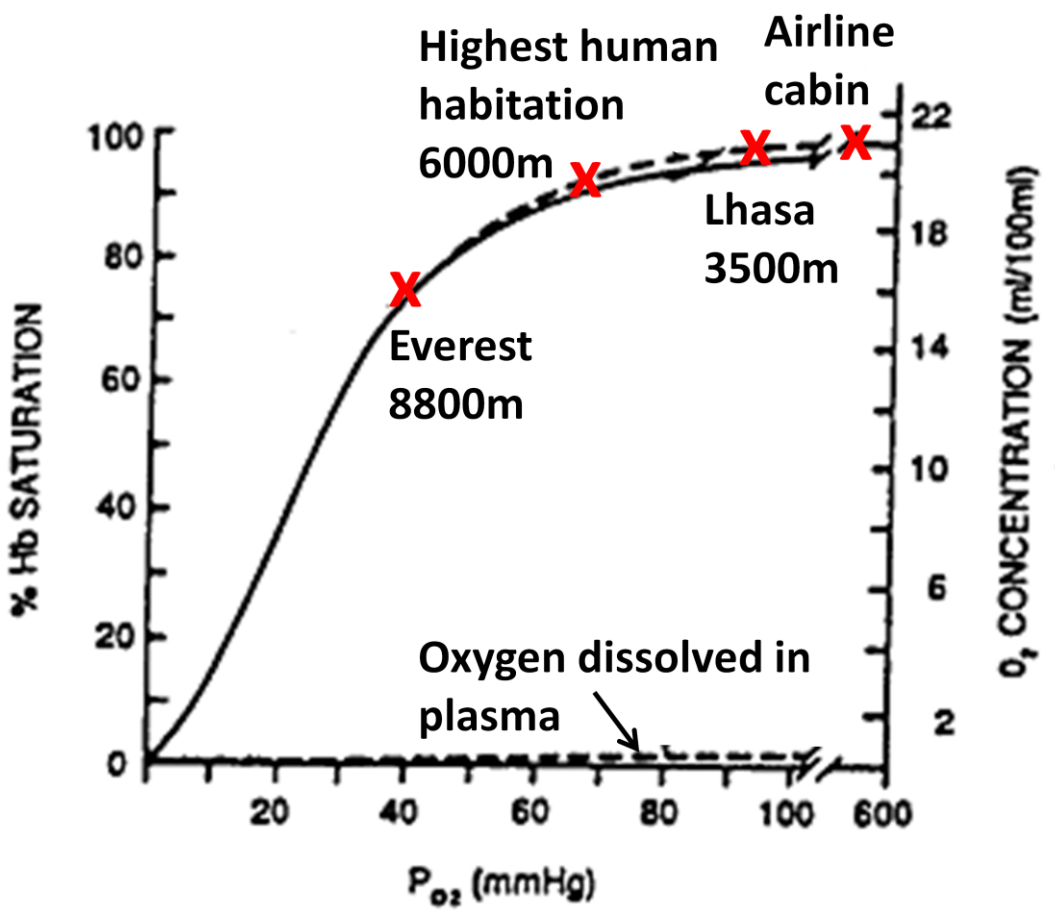


Figure 2.7-2 Oxygen-haemoglobin dissociation curve and hypoxia at altitude

Although ventilatory responses to hypoxic exposure show strong familial concordance (Fagan and Weil 2001) and familial clustering of HAHD and CMS has been observed, data is not complete enough to conclude that this does not result from shared environmental factors (Rupert and Koehle 2006). However, comparative studies of Tibetan, Han (lowland Chinese), Western and Andean populations have revealed significant differences in haematocrit (Beall and Reichsman 1984; Garruto, Chin et al. 2003) and pulmonary vascular musculature studied at matched altitudes (Gupta, Rao et al. 1992). These data are consistent with the hypothesis that genetic adaptation to a high altitude environment has occurred. In addition, pulmonary artery pressure of Aymara children at altitude is significantly lower than that of Caucasian children studied at the same altitude (Stuber, Sartori et al. 2008).

More robust evidence for genetic factors determining susceptibility to CMS comes from the observation that risk of CMS at a given altitude varies between different populations. Carlos Monge (who was the first to describe CMS) himself observed that Europeans were more susceptible to the disease than indigenous Andean residents (Monge 1943), and Tibetans, believed to have the longest high altitude antiquity, have strikingly reduced incidence of CMS compared with either Han Chinese or Andeans residing at the same altitude since birth (Pei, Chen et al. 1989; Moore 2001). This variation suggests that genetic modifiers of susceptibility to the disease may exist.

2.7.1.2 Reproduction at high altitude

That lowlanders have impaired ability to reproduce at altitude compared with indigenous high altitude populations has been recognised for generations: it is traditional for female Han Chinese residents of Tibet to stay with relatives in lowland China during pregnancy, returning soon after childbirth. Among the Han infants born at high altitude in Lhasa, Tibet (altitude 3658m above sea level) there is a higher incidence of subacute infantile mountain sickness (which consists of polycythaemia, pulmonary arterial hypertension and right heart failure) compared with Tibetan infants (Niermeyer, Yang et al. 1995). In addition, the birth weight and oxygen saturation of the Tibetan infants was significantly greater than those of the Han infants born at the same altitude. Other studies have shown that high altitude is an independent predictor of low birth weight (Keyes, Armaza et al. 2003) and that the susceptibility to this effect is different between different ethnic populations, with Andean and Tibetan populations displaying less reduction in birth weight at high altitude compared with European and Han Chinese (***Figure 2.7-3***) (Moore, Shriver et al. 2004). Early records show that when the Conquistadors (Spanish immigrants) arrived at high altitude regions of what is now Peru in the 16th Century, they experienced reduced fertility and increased perinatal mortality compared with the local Inca population (Gonzales 2007). This '*Curse on the Conquistadors*' is still evident today in Bolivia, where children born to Aymara parents (the indigenous group with the longest

high-altitude ancestry in the country), despite significantly more adverse socio-economic conditions, are protected against intra-uterine growth retardation and are consistently heavier at birth than children with European ancestry in whom poverty is less common (Julian, Galan et al. 2008). This suggests that even gross disparities of income cannot overcome either the effects of altitude on intrauterine growth or the genetic adaptation this stress has produced in high altitude communities (Giussani 2007; Bennett, Sain et al. 2008).

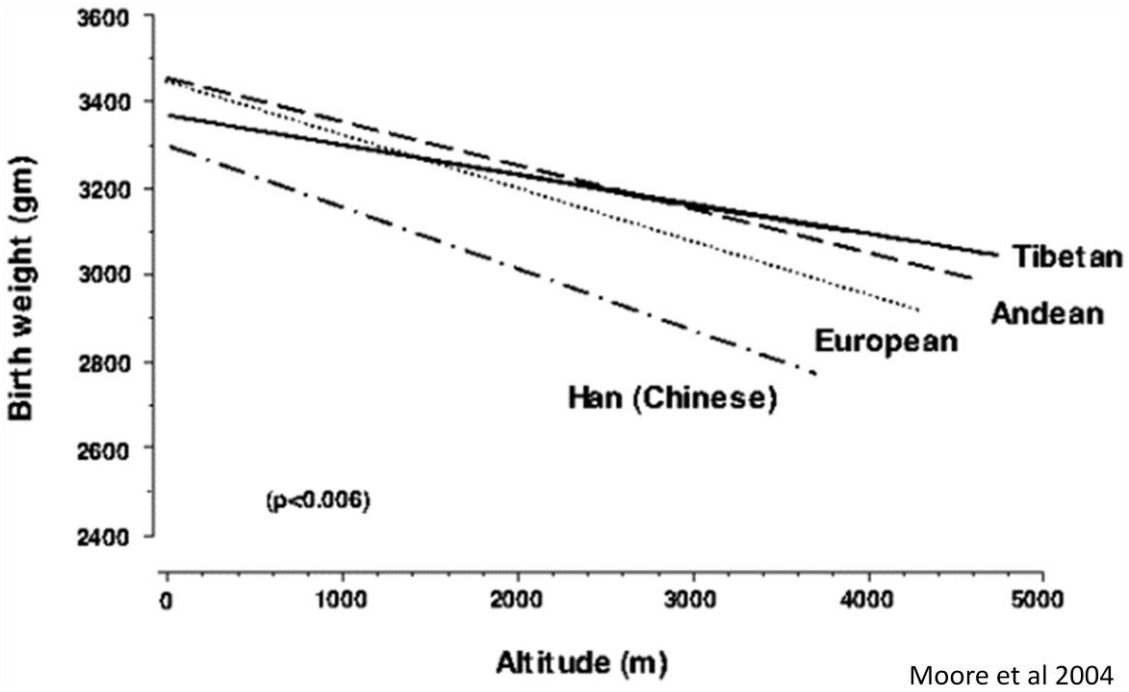


Figure 2.7-3 Birth weight at altitude by ethnicity
Altitude-associated reduction in birth weight is less pronounced in populations with high altitude ancestry (after Moore et al 2004).

Clearly, any genetic trait which has a direct effect on reproductive fitness will have a disproportionate propensity for selection, and populations in several parts of the world have lived at altitude for many generations, including Tibetans in the Himalayas, Aymara in the Andes and the Ambaras in the Ethiopian Highlands. Comparative studies of high altitude residents compared with lowlanders living at altitude have shown differences in various physiological parameters including lower haemoglobin levels. Importantly, these ancient high

altitude populations do not increase their haemoglobin levels at altitude. The Ambaras people of the Ethiopian highlands have a particular ability to maintain arterial oxygen saturation which is not understood, but is likely to reflect altered pulmonary physiology. The history of human settlement in these different regions likely explains why chronic mountain sickness is much more common in the Andes. Northern Ethiopia has been inhabited by human beings longer than almost everywhere else on Earth. Human habitation of the Himalayan plateaus is thought to date back >50,000 years and of the high Andes for significantly less, probably ~10,000 years. The Himalayan population has undergone significantly less admixture with lowland groups than other high altitude populations, primarily as a result of its isolation and inaccessibility – the Tibetan plateau has a diameter of over 750 km, lies 1600 km from the nearest ocean and is bounded by the world's highest mountains and one of its largest deserts. The high Andes, in contrast, is a narrow strip which lies within 200 km of the Pacific coastal plain and the Amazonian river system and has been subject to European colonisation for the last ~400 years.

2.7.1.3 Positive selection in Tibetans

While the case for a genetic adaptation for survival at high altitudes has seemed persuasive for a long time, it was not until 2010 that the first strong direct evidence for this effect (and the genes involved) was uncovered. Simonson et al used cross-population extended haplotype homozygosity (XP-EHH) analysis (Sabeti, Varilly et al. 2007) to identify alleles which have increased in frequency to the point of (near) fixation in the Tibetan population (compared with control populations from lowland East Asia in which there was variation at these loci). They also measured the extent of decay of linkage disequilibrium (the integrated haplotype score, iHS) between genomic locations and neighbouring regions in order to identify whether any variants are likely to have increased in frequency over a short enough time period to leave detectable LD across large distances, suggesting recent selection. While XP-EHH detects alleles which have reached or are approaching fixation in a population, iHS allows detection of alleles which have not yet reached fixation and can therefore be used to correlate genotype (or number of copies

of an allele) with phenotype within a population – a comparison which is robust to superimposed genetic differences between populations resulting from either genetic drift or selection at other loci. Starting with a panel of 247 *a priori* candidate genes (which were generated using Gene Ontology categories involving biological pathways implicated in the response to oxygenation), 10 of these, including *HIF2α* and *PHD2,* showed evidence of positive selection in the Tibetan population (Simonson, Yang et al. 2010). While *HIF2α* was only identified on the basis of XP-EHH, *PHD2* was identified by both XP-EHH and iHS. This allowed the researchers to show that the number of copies of the advantageous allele of *PHD2* was inversely correlated with haemoglobin level (each copy reducing the haemoglobin concentration by approximately 1.7 g/dl), consistent with a model in which *downregulation* of the HIF system is protective against the adverse effects of high altitude living. Since *HIF2α* was detected on the basis of fixation in the Tibetan community a similar quantitative trait analysis was not possible for the putatively advantageous allele of this gene, but it seems likely that if survival at altitude is enhanced by inheritance of an allele of *PHD2* which results in reduced haematocrit (presumably by enhancing HIF hydroxylation) then any advantageous *HIF2α* allele will have *reduced* stability compared with alleles common in the lowland population. Clearly, functional characterisation of the selected alleles will be needed to confirm this supposition.

It is notable that variation in HIF1α has not been found to be associated either with Mendelian traits in individuals with polycythaemia nor with positive selection in high altitude communities. While this may reflect a lack of an important role of this HIF subunit in environmental oxygen sensing and haematocrit regulation (i.e. it is rendered redundant in this aspect of oxygen sensing by HIF2α) it remains a possibility that there is such an absolute dependence on HIF1α for the normal operation of the oxygen sensing pathway in humans that any functionally significant change in the protein is incompatible with survival to reproduction.

2.8 Conclusion

There is now evidence from independent studies of Mendelian human traits, human populations, genetically modified model organisms and *in vitro* systems which implicates HIF2α as a central component of environmental oxygen sensing in humans. Understanding the mechanisms underlying the biological effects of hypoxic exposure, its physiological responses and genetic adaptations, is important not only for physiologists, mountaineers, aviators and tourists, but also for physicians: in addition to chronic mountain sickness, many disorders, such as cyanotic heart disease, chronic pulmonary disease and obstructive sleep apnoea, are associated with prolonged systemic hypoxia. Furthermore, there is evidence that hypoxia plays a major role in the aetiology and pathogenesis of pre-eclampsia (Gilbert, Ryan et al. 2008) and intrauterine growth restriction (IUGR) which are major causes of pre- and perinatal mortality worldwide. Although no therapies currently exist which are able to target selectively specific branches of the HIF pathway, it seems logical to predict that significant benefit to large numbers of people would accrue from such treatments if they were developed in the future.

Chapter 3: Identification of CFHR5 nephropathy

3.1 Introduction: Renal disease and Complement

This investigation was undertaken in order to elucidate the cause of an unusual form of glomerulonephritis occurring in multiple members of a British Cypriot kindred living in London. Consideration of the pathology and phenotype suggested that dysregulation of the complement alternative pathway was responsible for the disease and this observation informed a combined linkage and candidate gene approach which was used to identify a mutation in the *Complement Factor H Related protein 5* gene which cosegregated with the disease.

Glomerulonephritis is a leading cause of kidney disease worldwide and can be subdivided on the basis of histological appearance or by clinicopathological features. While some glomerulonephritides are most often found as sporadic diseases of the kidney, others are now appreciated to represent renal manifestations of systemic diseases, which can have genetic or environmental aetiologies.

Mesangiocapillary glomerulonephritis (MCGN, or alternatively membranoproliferative glomerulonephritis, MPGN) refers to the light microscopic appearances of glomerular mesangial cellular proliferation with thickening of the capillary walls and was observed in affected members of the family under investigation. MCGN can occur as a consequence of a number of disease processes, but central to its pathophysiology is the deposition of complement-containing electron dense material in the glomerulus.

While MCGN can be subdivided in a number of ways on the basis of light or electron microscopic appearance, it can also be categorized on the basis of aetiology. In some individuals, MCGN occurs as a consequence of systemic inflammation resulting from chronic infection (for example endocarditis, hepatitis or osteomyelitis) or owing to autoimmune conditions (such as systemic lupus erythematosus, SLE). In these conditions immunostaining of the glomeruli reveals the presence of immunoglobulins as well as complement in the electron

dense material and it is likely that it is the immune response and associated immunoglobulin production which drives the renal injury in this context. Why some individuals with these inflammatory disorders develop MCGN and others do not is poorly understood.

Rarely, MCGN may occur as a result of dysregulation of complement activation in the absence of an antibody-driven inflammatory process. In this situation, immunostaining reveals complement but not immunoglobulin deposition within the glomerulus. Recently, it has been recognised that these conditions can be divided into dense deposit disease (DDD, or alternatively MCGN type 2), in which there is linear dense transformation of the glomerular basement membrane, and C3 glomerulonephritis (C3GN) in which there is no dense transformation of the glomerular basement membrane but there is C3 deposition in the mesangium and along the capillary wall in the absence of immunoglobulins. This can be seen with or without light microscopic features of MCGN (for review see (Pickering and Cook 2008)). Examination of renal biopsy specimens from affected individuals from the family under investigation revealed features compatible with C3GN.

In addition to MCGN, disorders of complement activation can also cause the Haemolytic Uraemic Syndrome (HUS), in which the pathological appearances are of thickening of arterioles and capillaries, endothelial swelling and accumulation of protein and cellular debris in the widened subendothelial space. In addition there is intravascular haemolysis (with erythrocyte fragments visible on a peripheral blood film) with platelet thrombi occluding arterioles. These histological features are typically seen in the kidney, although other organs (especially the heart and brain) may be affected. HUS may occur as a consequence of exposure to Shiga-like toxin which is produced by several *Escherichia coli* serotypes (particularly O157) and by *Streptococcus pneumoniae,* and the histological lesion is indistinguishable from non-Shiga-toxin associated (or atypical) HUS (aHUS), in which host abnormalities in complement regulation have frequently been identified.

Understanding the aetiology and pathogenesis of these conditions has aided understanding of the complement system and it may be that, in the future, diagnostic entities based on pathological morphology will give still more ground to those defined by their molecular pathophysiology.

3.2 Complement

Complement is a key component of the immune system which underlies the long-recognized ability of serum to "complement" antibodies in their ability to destroy micro-organisms (von Fodor 1887; Bordet 1895). Complement as a distinct substance was first identified in 1901 by Bordet and Gengou who showed that antibody-antigen complexes removed the ability of serum to lyse antibody-coated cells – work for which Bordet received the 1919 Nobel Prize in Physiology or Medicine. Subsequent discoveries, over the last century or more, have revealed that complement is composed of a large number of predominantly circulating proteins (with a combined concentration of over 3g per litre in plasma), the central components of which are proteases.

It is now recognised that the complement system has three primary physiological activities: defence against pyogenic bacteria; forming an interface between innate and adaptive immunity; and disposing of immune complexes and other products of inflammation (Walport 2001). The nomenclature of the complement system reflects the order of discovery of each component, not the order in which they are thought to have arisen over the course of evolution or the order in which they interact biochemically. That consideration aside, since the advent of molecular biology and its interface with clinical genetics, complement biologists have unlocked the secrets of an elegant, important and highly conserved molecular apparatus which is central to the immune system.

Since microbial attack is potentially lethal and self promoting (once established, microbes are able to reproduce, increasing the resources required by the host to clear them) it is important that, once a microbial invasion is recognised, the host is able to mount a response of appropriate magnitude as quickly as possible. This requirement probably underlies the property of activated complement components (particularly C3) to catalyse their own activation, providing a positive feedback loop which exponentially amplifies a danger signal. This property is shared by other facets of innate immunity such as macrophage activation (for review see (Hu, Chakravarty et al. 2008)). Clearly, any system which is both powerful (in terms of its ability to destroy cells and organisms) and autocatalytic requires finely tuned regulatory mechanisms in order to prevent inappropriate or excessive activation and to terminate the activation once the stimulus has been removed. For this reason, defects of complement regulation can have dire consequences for the host and contribute significantly to the burden of human disease.

The complement cascade can be activated in a variety of ways, conventionally considered as comprising 3 distinct pathways, known as the classical, mannose-binding lectin and alternative pathways, which all converge on (i.e. result in) the cleavage of C3 into C3a and C3b which, in turn, results in the activation of a common terminal complement pathway.

3.3 The terminal pathway

Activation by cleavage of the terminal complement component C5 produces the C5b and the anaphylatoxin C5a. C5a, a 10kDa cleavage fragment, is able to bind with high affinity to receptors (termed C5aR and C5L2) on the surface of inflammatory cells, activating and recruiting them to sites of complement activation (Peng, Li et al. 2009). In common with other anaphylatoxins, the effects of C5a are protean, involving alterations of apoptosis, lipid metabolism, innate immune responses and adaptive immune responses through effects on antigen-presenting cells and T cells (Klos, Tenner et al. 2009). The larger fragment, C5b, initiates

formation of the tubular Membrane Attack Complex (MAC) by recruitment of C6, C7, C8 and polymerised C9. The MAC is able to insert into the surface of a targeted cell leading to its lysis, a process which is regulated by CD59, a glycosyl-phosphatidylinositol (GPI)-anchored protein ubiquitously expressed on host surfaces (Meri, Waldmann et al. 1991; Davies and Morgan 1993). The importance of this regulatory protein is indicated by the observation that its deficiency (which may be inherited or acquired) results in complement-mediated lysis of host erythrocytes and Paroxysmal Nocturnal Haemoglobinuria (PNH) (Yamashina, Ueda et al. 1990).

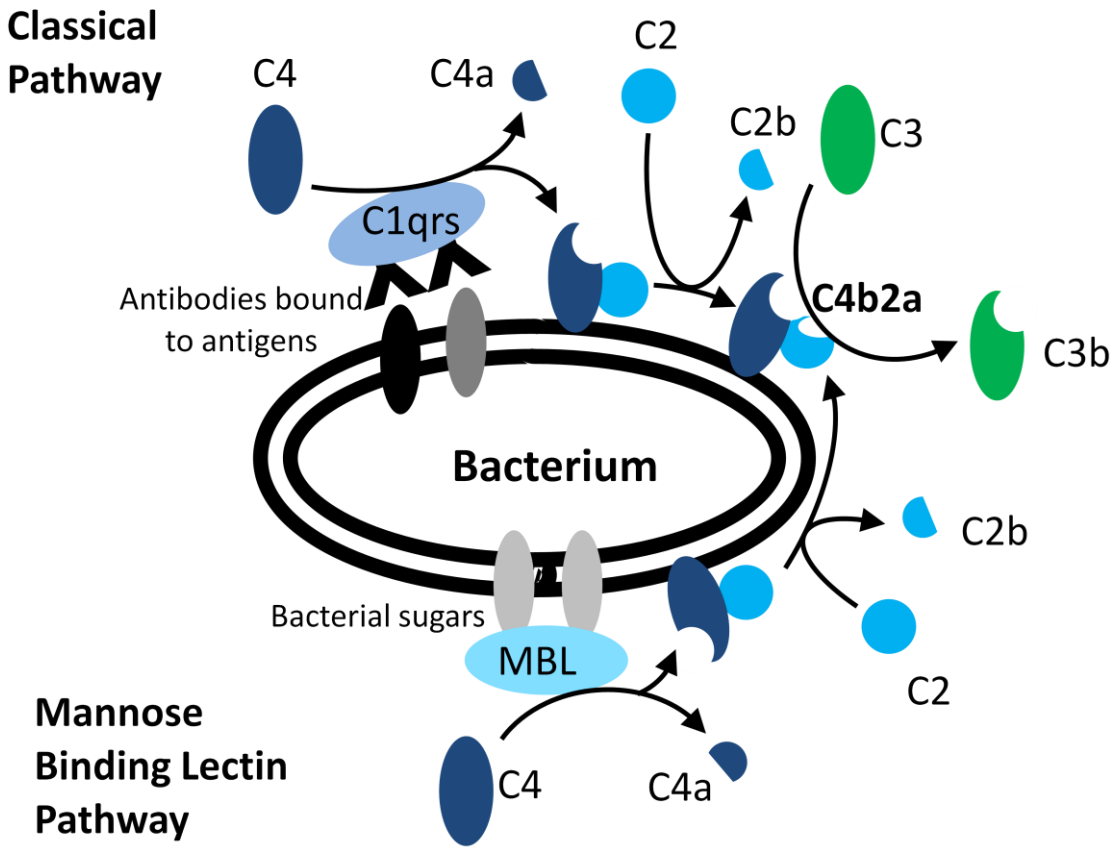


Figure 3.3-1 The Classical and Mannose Binding Lectin complement pathways

3.4 The classical pathway

The classical pathway was the first to be discovered (hence the name) and in fact underlies many of the initial 19th century observations which resulted in the discovery of complement system. Initiation of the classical pathway results from binding of the C1 complex (which is itself composed of the proteins C1q, C1r and C1s in the ratio 6:2:2) to antibodies bound to bacterial surfaces. The C1rs heterotetramer has catalytic activity which results in cleavage of C4 in the circulation to produce C4b (and C4a). While C4a is an anaphylatoxin which remains in solution and acts as a signal to recruit inflammatory cells, C4b becomes covalently bound to the cell surface and is able to bind C2. C4b bound C2 is then cleaved by C1s within the activated C1 complex, resulting in the formation of C4b2a which is covalently anchored to the cell and itself has catalytic activity resulting in cleavage of C3 to produce C3b and the anaphylatoxin C3a (***Figure 3.3-1***). The C4b2a complex is also termed the C3 convertase of the classical pathway and, when bound to C3b (to produce C4b2a3b), becomes the C5 convertase of the classical pathway which cleaves C5 to activate the terminal pathway. The absence of immunostaining for C1q in biopsies from family members under investigation here suggested that the genetic defect did not result in antibody or classical pathway deposition in the glomerulus and therefore did not suggest a defect of classical pathway activity.

3.5 The mannose-binding lectin pathway

The mannose-binding lectin (MBL) pathway is activated by the formation of a complex composed of MBL and MBL-associated proteases 1 and 2 (MASP1 and MASP2). The formation of this complex is triggered by mannose groups which are present on bacterial cell surfaces. The activated MASP2 homodimer (in a similar fashion to the C1rs heterotetramer, see ***Figure 3.3-1***) is able to cleave C4 and C2 to produce the covalently bound C4b2a C3 convertase, although it is now believed that MASPs are promiscuous, with protease activity for substrates outwith the recognised complement system (Hajela, Kojima et al. 2002; Dobo, Harmat et al.

2009). A key player in the regulation of both the classical and MBL pathways is C1 inhibitor which inhibits the activities of a range of soluble proteases, including C1r, C1s, MASPs and proteases of the coagulation cascade (Davis 2004; Wallis, Mitchell et al. 2010). The importance of this regulatory mechanism is illustrated by the clinical consequences of heterozygous genetic deficiency (or acquired loss) of this protein, resulting in life-threatening episodes of angio-oedema, urticaria, gastro-intestinal mucosal oedema and kinin-mediated capillary leakage.

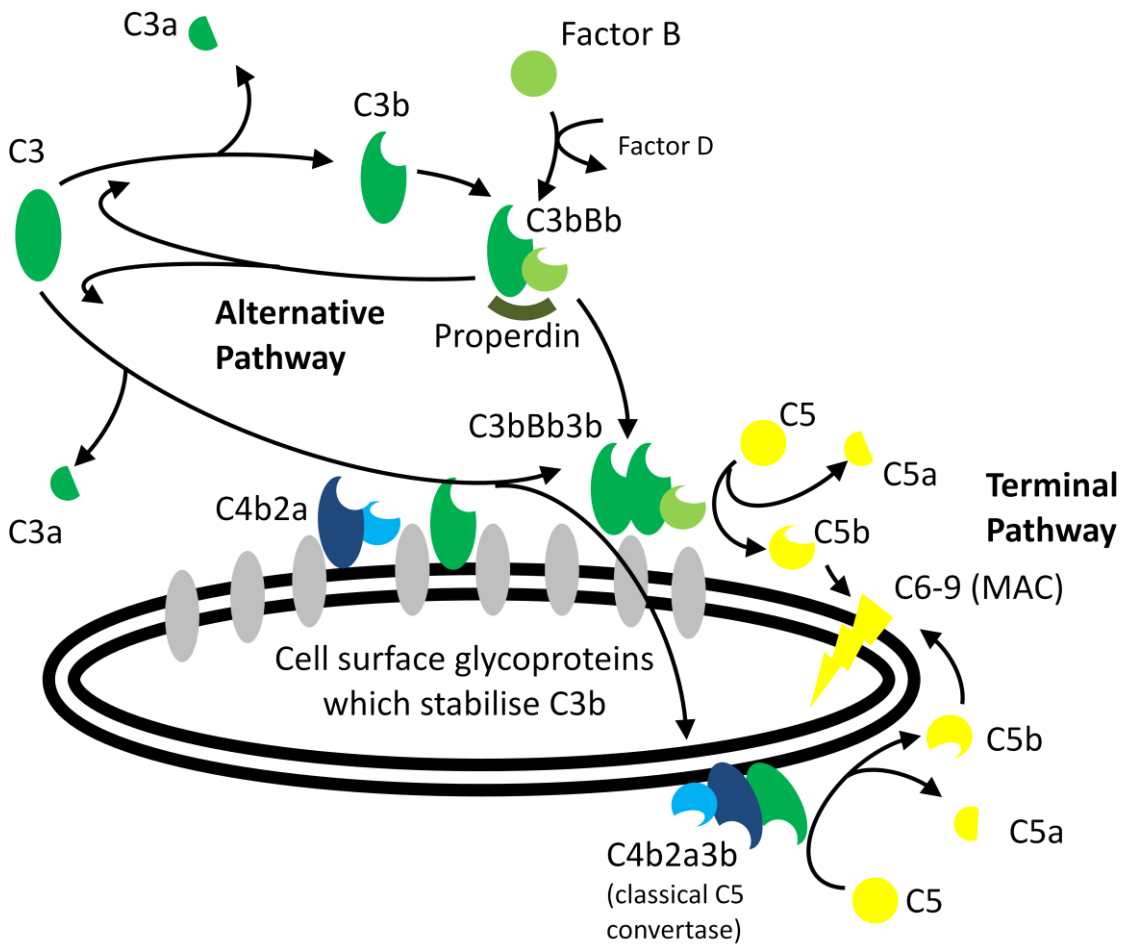


Figure 3.5-1 Activation of the Alternative and Terminal Complement pathways

3.6 The alternative pathway

Central to the alternative pathway (AP) is C3b (see ***Figure 3.5-1***) which is continuously generated by low-grade cleavage of C3 (by water hydrolysis (Pangburn, Schreiber et al. 1981)) in the plasma – a process known as AP tickover. C3b is able to bind (and is stabilised by)

hydroxyl groups of cell-surface proteins and carbohydrates as well as Factor B which can then be cleaved by factor D to form C3bBb, also known as the C3 convertase of the AP. C3bBb catalyses the cleavage of C3 to form more C3b and the anaphylatoxin C3a which plays an important role in the recruitment of inflammatory cells to sites of C3 activation. C3bBb is stabilized by properdin and can, in addition, bind free C3b (to form C3bBb3b) which is the C5 convertase of the AP which activates the terminal pathway. Since these complexes are non-covalently bound they also dissociate freely, and certain factors (discussed below) can enzymatically accelerate this dissociation – a property known as 'decay acceleration'. Clearly, stabilization of C3b or the C3bBb complex (for instance by the presence of an appropriate 'protected' surface, such as a bacterial cell) will result in secondary generation of more C3b in a positive feedback loop until either C3 is depleted or the loop is terminated by the inactivation of C3b (see below). Furthermore, since cleavage of C3 to form C3b is a consequence of activation of both the classical and MBL pathways, activation of either of these pathways results in AP activation and hence amplification of the signal. Consequently it can be argued that the AP, rather than being a third distinct complement activation pathway, could best be regarded as the downstream amplification limb of the other 2 pathways which also has autonomous activity itself, rather like the petrol supply to a car engine which ticks over without any intervention from the driver but can be increased by depressing the accelerator pedal. The observation that, in the family under investigation, metabolites of C3 in the absence of C1q or immunoglobulins were seen in renal biopsies of affected individuals suggested that increased C3 convertase activity was contributing to the disease – most likely through aberrant activation of the alternative pathway.

3.7 Other mechanisms of AP convertase activation

In addition to the mechanisms described above, physiological generation of the AP convertase (C3b) can also be effected by other enzymes, including microbial proteases,

elastase, which is released by neutrophils and normally inhibited by alpha-1-antitrypsin (Johnson, Ohlsson et al. 1976), and coagulation factors (see below). Pathological stabilisation of the AP convertase can result from the acquisition of an IgG antibody (termed C3 Nephritic Factor or C3NeF) which binds to C3bBb, preventing its dissociation or cleavage by other components of the complement cascade. This results in uncontrolled C3 consumption and fluid phase and surface-bound complement activation leading to dense deposit disease (DDD), and C3NeF is the commonest cause of DDD worldwide (Licht and Fremeaux-Bacchi 2009). DDD in this setting is also associated with the deposition of a substance, morphologically similar to the dense material found in the glomerular basement membrane, in Bruch's membrane of the eye leading to the fundoscopic appearance of drusen (Duvall-Young, MacDonald et al. 1989). It is now widely believed that drusen in DDD represent an ocular manifestation of the same pathophysiological process as that responsible for the renal disease – namely dysregulated AP activation. The question of whether there is a consequently increased risk of visual impairment in DDD has not yet been conclusively answered (Colville, Guymer et al. 2003; D'Souza, Short et al. 2008). Drusen are also the clinical hallmark of age-related macular degeneration (AMD) and determining whether factors which affect systemic complement activation among the general population can also modify the risk of AMD is an area of active research.

Activating mutations in the genes for C3 and Factor B which lead to stabilisation of the C3 convertase have also been described in association with aHUS (Goicoechea de Jorge, Harris et al. 2007; Fremeaux-Bacchi, Miller et al. 2008), demonstrating the importance of the normal function of these proteins for complement regulation in humans. Intriguingly, mutations resulting in haploinsufficiency (i.e. deficiency) of secreted C3 have also been described in association with aHUS, findings not well explained by the current paradigm (Fremeaux-Bacchi, Miller et al. 2008).

The ability of such a wide range of enzymes and antibodies to activate C3 implied that the molecular defect responsible for the family under investigation here could reside in one of a number of genes – from either within or beyond the recognised complement system.

3.8 Breakdown products of C3b

As described above, C3 is cleaved to form C3b and the anaphylatoxin C3a. In addition to contributing to the AP convertase, C3b is able to bind to biological surfaces due to the exposure of an internal thiolester bond. This allows C3 convertase and C5 convertase activity to be maximised at a bacterial surface. C3b (whether bound or unbound) can be modified by the action of an array of AP regulating proteins (discussed below) first into iC3b (by cleavage between residues 1281 and 1282, and residues 1298 and 1299) and subsequently into C3c and C3dg (by cleavage between residues 932 and 933) (Davis, Harrison et al. 1984) and C3dg is cleaved by multiple proteases (including trypsin) to form C3d which can remain covalently bound to the surface.

iC3b and its breakdown products do not have convertase activity (and therefore are not able to contribute to amplification of the complement cascade) but are able to act as opsonins. In addition, iC3b has pro-inflammatory effects (mediated primarily by its ability to act as a ligand for complement receptors on phagocytic cells) which are not all shared by C3dg and C3c.

There are at least 5 complement receptors (CRs) described to date, termed CR1-4 and CRIg. CR1 and CR2 are glycoproteins which contain multiple homologous short consensus repeat domains and are products of homologous genes situated near each other on Chromosome 1. CR1 acts predominantly as an inhibitory regulator of complement activation and is discussed in detail below. CR2 is present on B cells and is the principal CR which, following ligand binding, enhances B cell immunity (Weis, Tedder et al. 1984) by increasing B cell receptor signalling and promoting survival (Matsumoto, Kopicky-Burd et al. 1991). It is able to bind iC3b, C3d and

Interferon-α resulting in an inflammatory pattern of gene expression (Clemenza and Isenman 2000; Asokan, Hua et al. 2006). Allelic variants of CR2 are associated with altered risk of SLE (Wu, Boackle et al. 2007). In addition, CR2 is bound by Epstein Barr Virus during infection (Yefenof, Klein et al. 1976).

CR3 and CR4 are members of the β_2-integrin superfamily. CR3, also termed CD11b, macrophage-1 antigen (MAC-1) or integrin alpha M (ITGAM), is expressed by leukocytes involved in the innate immune system, including monocytes, granulocytes, macrophages, and natural killer cells. CR3 stimulation by iC3b contributes to leukocyte adhesion and activation of these cells at sites where complement is activated (Solovjov, Pluskota et al. 2005). C3dg is unable to stimulate CR3, although cleavage of iC3b by kallikrein (which occurs between residues 923 and 924) yields a fragment, known as C3d-k, which is able to stimulate CR3. It has also been shown that the synthetic 9-amino acid peptide corresponding to residues 924-932 of C3 is sufficient to induce leukocytosis in experimental systems (Hoeprich, Dahinden et al. 1985).

CR4 (also known as CD11c or alternatively ITGAX) is less well studied but has been shown to be present on the surface of neutrophils and platelets where it binds iC3b and C3dg with high affinity (Vik and Fearon 1987). CR4 has also been identified on the surface of dendritic cells where it plays a critical role in adhesion-mediated phagocytosis following opsonisation by activated C3 (Ben Nasr, Haithcoat et al. 2006). It is notable that polymorphic variation at the genetic locus of CR4 is also associated with altered risk of systemic lupus erythematosus (SLE) (Hom, Graham et al. 2008).

CRIg is encoded by a gene on the X chromosome and its expression is limited to certain tissue macrophages found in the Kupffer cells, heart, synovial lining, alveoli, placenta, adrenal gland and foam cells in atherosclerotic plaques (Helmy, Katschke et al. 2006). CRIg has been shown to bind the β chain of C3 degradation products and is thought to play a significant role in clearing opsonised pathogens and debris from the circulation (Bilzer, Roggel et al. 2006).

C3 and its cleavage products therefore have an array of properties ranging from primed (uncleaved C3), autocatalytic (C3b), opsonising (C3b, iC3b, C3c and C3dg) and inflammatory (C3a, iC3b, C3d-k).

3.9 Evolution of the AP

The AP is present in all vertebrates which have been studied and is probably the most ancient of the three pathways since, while homologues of the classical pathway and MASP genes have to date been identified only in chordates, genes with homology to AP components C3 and Factor B have been identified in the sea anemone *Nematostella vectensis* (Nonaka and Kimura 2006) which belongs to the phylum Cnidaria. The Cnidaria share a common ancestor with Bilateria not less than 1,300 million years ago, before the divergence of the Protostomia (which includes arthropods and nematodes) from the Deuterostomia superphylum (which includes the chordates), thought to have occurred approximately 1,000 million years ago (Pinto, Melillo et al. 2007). The absence of complement pathway genes in protostome species (including *Caenorhabditis elegans*) has been suggested to be a consequence of secondary loss which has occurred numerous times (Nonaka and Kimura 2006), and this theory is supported by the detection of a complement-like system in the protostomes *Drosophila melanogaster* (Lagueux, Perrodou et al. 2000) and *Anopheles gambiae* (Levashina, Moita et al. 2001).

It is a commonly held view that the complement and the coagulation cascades share a common ancestral set of genes and indeed there is ample evidence of cross-talk between the two systems: coagulation factor (F) XIIa is able to activate C1q ; Kallikrein can cleave iC3b, albeit at a site 9 residues N-terminal to the canonical Factor I mediated cleavage site, to produce C3d-k (Meuth, Morgan et al. 1983); C1 inhibitor can inactivate Kallikrein, FXIIa, FXIa, plasmin, t-PA (Davis 2004; Wallis, Mitchell et al. 2010); FXa, FXIa, thrombin and plasmin can cleave C5 and C3 to produce active C5b and C3b and anaphylatoxins C5a and C3a (Amara, Rittirsch et al. 2008)

and in addition mast cells exposed to C5a are induced to switch from producing the fibrinolytic t-PA to the pro-thrombotic PAI-1 (Wojta, Huber et al. 2003). It is not completely clear to what extent this cross-talk represents physiologically important adaptive mechanisms which are beneficial (in an evolutionary sense) or alternatively maladaptive (or artefactual) consequences of 'accidental' (or experimental) activation of one pathway by the extreme activation of the other owing to the homology between proteins of both pathways. Indeed, many of the most serious consequences of bacterial infection, such as septic shock and disseminated intravascular coagulopathy, may arise from such cross-talk and would seem, at least in some settings, to be detrimental to the host organism.

There is evidence that genetic variation in humans resulting in 'deficiency' of complement proteins (potentially resulting in reduced ability to clear microorganisms) could provide a survival advantage in some circumstances – particularly those in which microorganisms have themselves subverted components of the host complement system to increase their virulence. One example of this is the observation that reduced levels of MBL in Ethiopians appears to confer protection against lepromatous leprosy, potentially by reducing the ability of the mycobacteria to gain entry into macrophages (Garred, Harboe et al. 1994; Schorey, Carroll et al. 1997; Walport 2001). There is therefore mounting evidence that, in common with adaptive immunity, the innate immune system is dynamic and responsive (in an evolutionary sense) to environmental influences and pathogens. An intriguing question is whether genetic 'defects' which result in *increased* complement activity could provide enhanced innate immunity – a situation which may have provided a survival advantage among children infected with bacteria (in whom adaptive immunity is less well developed (Lee, Andalibi et al. 2004)) in the pre-antibiotic era.

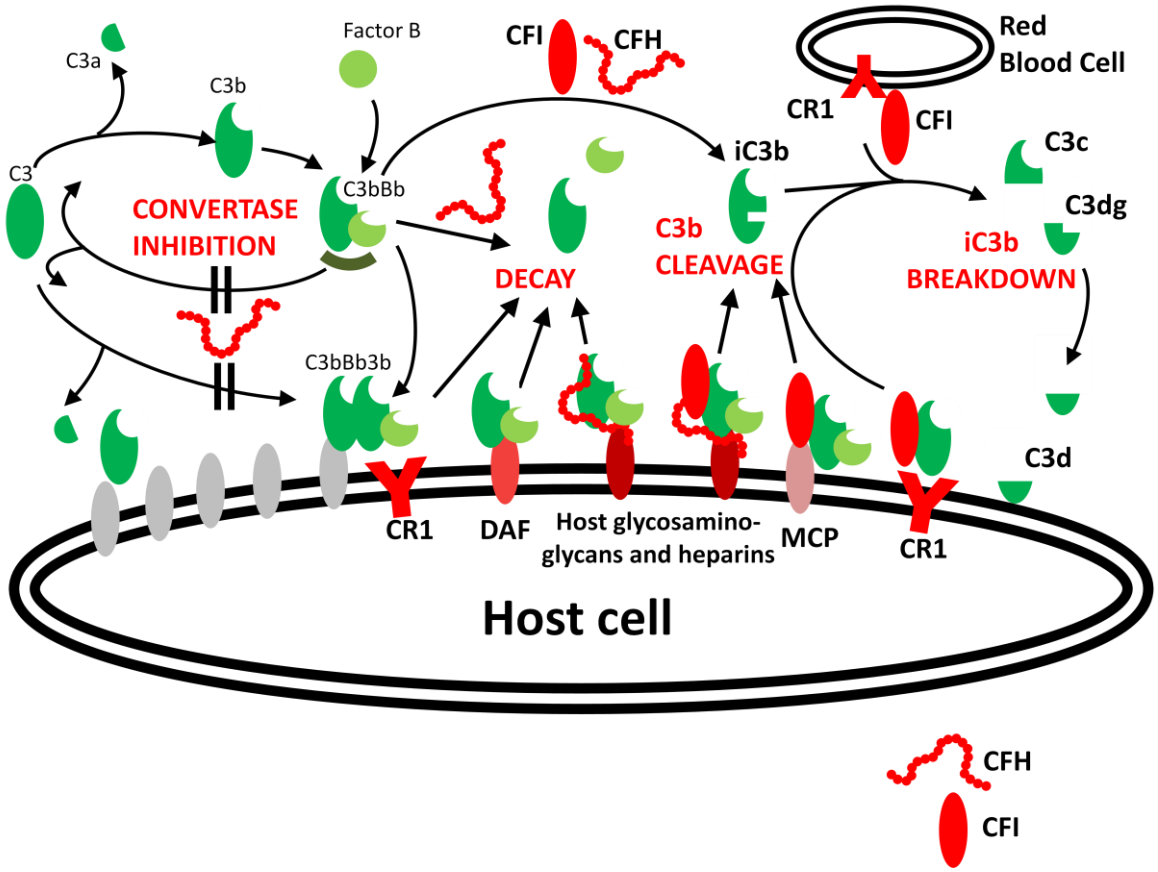


Figure 3.9-1 Regulation of the Alternative Complement Pathway

3.10 Modulation of alternative pathway activation

As discussed above, the alternative pathway contains significant capacity for autocatalysis and runaway activation since C3b can bind to factor B and subsequently cleave C3 to generate more C3b. In this context, it is clear that regulatory mechanisms must exist to prevent runaway activation which could result in complement-mediated damage to host cells or organs (for example as in C3GN) and secondary depletion of C3 leading to impaired ability to destroy invading micro-organisms. Conceptually, the AP can be considered to consist of 2 cycles – an activation cycle (described above) and a regulation cycle, both of which are active in both the circulation and on surfaces (***Figure 3.9-1***).

Peter Lachmann has proposed the attractive hypothesis that the development of the molecular apparatus able to downregulate AP activity became imperative with the evolution of

a pumped circulation, since in smaller organisms (in which diffusion is sufficient to provide oxygen and nutrients to tissues) any areas of complement activation will be localised. In addition, larger organisms also produce a large range of abundant circulating proteases which may be capable of cleaving C3 aberrantly (Lachmann 2009). These factors have probably contributed to the evolution of genes coding for a number of proteins which regulate AP activity on host surfaces, including Complement Receptor 1 (CR1 or CD35); Decay Accelerating Factor (DAF or CD55); Membrane Cofactor Protein (MCP or CD46). In addition the circulating proteins Complement Factor I (CFI) and Complement Factor H (CFH) are known to play crucial roles in the regulation of C3 turnover in the circulation as well as on surfaces.

3.10.1 The 'protected surface'

Complement activation is enhanced at biological surfaces – examples being bacterial cells and desialated red cells (Kazatchkine, Fearon et al. 1979). This phenomenon is known as the 'protected surface' which is a rather misleading term since it is complement activation which is 'protected' resulting in the surface itself being at *increased* susceptibility to complement-mediated damage (i.e. not at all protected!) The biochemical mechanisms underlying the protected surface are not well understood but probably include the ability of complement components (such as C3b and C4b) to form covalent bonds with hydroxyl groups of cell surface proteins. It is likely that the consequent need to shield host surfaces from complement-mediated self-attack has driven the evolution of genes for the regulatory proteins which are expressed on, or preferentially bind to, host surfaces and tip the balance from activation to regulation of the AP. Clearly, failure of complement regulation at a host 'protected surface' can result in disease. An example of this is aHUS in which genetic defects of complement regulation at host surfaces have frequently been identified. The deposition of C3 on the glomerular basement membrane in the absence of depletion of circulating C3 in affected members of the

family under investigation suggested that a specific impairment of complement regulation at host surfaces could be responsible for the disease.

3.10.2 Complement Receptor 1

Complement Receptor 1 (CR1 or CD35) is a 190 kDa transmembrane glycoprotein expressed on the surface of most blood cells (including erythrocytes). CR1 has several functions: firstly, it acts as a cofactor for the cleavage of C3b and also iC3b by Complement Factor I; secondly it can accelerate the decay of the C3 and C5 convertases (both classical and alternative); and thirdly, CR1 is an immune adherence receptor which can bind to C3b and C4b, allowing blood cells to bind to serum-exposed foreign particles (Ahearn and Fearon 1989). The consequences of this depend on the type of blood cell: CR1 on red cells results in transfer of the opsonised antigen to the spleen or liver where they can be internalized and degraded by macrophages; CR1 on circulating monocytes/macrophages allows immediate phagocytosis and degradation of the antigen; and CR1 on B lymphocytes and dendritic cells enhances localization and presentation of complement-coated antigens to the cells of the adaptive immune system.

The gene for CR1 lies on the long arm of Chromosome 1 (at 1q32) within a group of complement regulatory genes (including CR2) termed the Regulators of Complement Activation gene cluster (RCA) (Rodriguez de Cordoba, Lublin et al. 1985). In common with other members of the RCA, CR1 contains homologous short consensus repeat (SCR) domains which, in CR1, are critical for C3b and C4b binding, decay accelerating and cofactor functions (Smith, Mallin et al. 2002). SCRs consist of 60-70 amino acids and are arranged head-to-tail within complement regulatory proteins, usually separated by between 4 and 8 (non-conserved) residues. SCRs can vary greatly in sequence but each contains 4 conserved cysteines (producing 2 disulphide binds) and a conserved tryptophan. They have a hydrophobic core surrounded by beta sheet and each SCR folds independently. SCRs can have a number of biochemical functions, including

substrate binding and accelerating decay of complement complexes. In addition, they can have cofactor activity for Complement Factor I-mediated cleavage of C3b or iC3b.

Genetic variants of CR1 have not (to date) been implicated in any Mendelian disorders affecting humans, but there is evidence that variation in the gene can confer altered susceptibility to infectious diseases. In severe malaria, erythrocytes infected with *Plasmodium falciparum* can bind to uninfected red cells (forming the appearance of 'rosettes' on a blood film). CR1 plays a role in the formation of these rosettes by binding to a parasite-derived membrane protein and CR1 polymorphisms which are common in Africans are associated with reduced rosetting and have been suggested to improve resistance to severe disease (Rowe, Moulds et al. 1997). Intriguingly, one of these alleles (termed S12 in the Knops blood group typing system), while associated with reduced susceptibility to severe falciparum malaria, is associated with *increased* susceptibility to chronic *Madurella mycetomatis* infection, which causes mycetoma and is endemic in Sudan (van de Sande, Fahal et al. 2007). In addition, genetic variants resulting in deficiency of CR1 on the surface of erythrocytes have been identified with very high prevalence (up to 80%) in regions of Papua New Guinea in which falciparum malaria is endemic (Cockburn, Mackinnon et al. 2004). This is therefore potentially another example of environmental exposure producing genetic variation in complement activation in humans.

3.10.3 Decay Accelerating Factor

Decay Accelerating Factor (DAF or CD55) is a 70kDa glycoprotein composed of 4 consecutive SCRs tethered to the membrane by a glycosylphosphatidylinositol anchor (Kim and Song 2006). DAF is coded for by a gene which lies within the RCA at 1q32, allelic variants of which have been associated with altered risk of allergic respiratory diseases (Kawai, Takeshita et al. 2009). DAF is expressed by a large number of tissues in humans, acts to inhibit the formation of C3 convertases and accelerates their decay (Miwa and Song 2001).

3.10.4 Membrane Cofactor Protein

Membrane Cofactor Protein (MCP or CD46) is expressed by a large number of tissues in humans. It is also a member of the RCA cluster and consists of 4 SCRs bound to the membrane by a transmembrane domain. MCP has the ability to act as a cofactor for Complement Factor I-mediated cleavage of C3b and C4b (Lublin, Liszewski et al. 1988), thereby contributing to the regulation of complement at host 'protected' surfaces which expresses it. In addition, MCP can act as the cellular receptor for a number of pathogenic organisms, including measles (Dorig, Marcil et al. 1993), adenoviruses (Gaggar, Shayakhmetov et al. 2003), human herpesvirus-6 (Santoro, Kennedy et al. 1999) and Neisseria species (Kallstrom, Blackmer Gill et al. 2001). It has also been demonstrated that antibody-mediated stimulation of MCP can induce regulatory T cells to proliferate and release IL10, establishing a potential mechanism by which MCP links innate to adaptive immunity (Kemper, Chan et al. 2003). Mutations in the MCP gene have been implicated in atypical Haemolytic Uraemic Syndrome (aHUS) (Noris, Brioschi et al. 2003; Richards, Kemp et al. 2003; Caprioli, Noris et al. 2006). Whether there is an environmental context in which variants of MCP confer decreased susceptibility to infection is not known. Importantly, there is a low risk of recurrent renal disease in individuals who have received a renal transplant, presumably because the transplanted kidney endothelium predominantly expresses normal (i.e. donor-derived) MCP.

3.10.5 Complement Factor I

CFI is an 'induced fit' enzyme of 88 kDa circulating with a plasma concentration of approximately 35 µg/ml (de Paula, Barbosa et al. 2003). CFI requires the presence of a cofactor to catalyse the conversion (by cleavage at 2 sites in the C3b α-chain) of C3b to iC3b which is unable to bind to factor B and hence is unable to act as a C3 convertase. In addition, CFI can catalyse the further degradation of iC3b (by cleavage between residues 932 and 933) into the inactive breakdown products C3c, which is soluble, and C3dg, which can remain surface-bound. This reaction is accelerated efficiently by CR1 but not by CFH (Ross, Lambris et al. 1982).

CFI is coded for by a gene located at Chromosome 4q25 and homozygous deficiency of CFI in humans has been described. This is associated with depletion of complement C3 and Factor B in the circulation (in fact almost all the C3 present is detected as C3b), reduced AP activity and increased susceptibility to infections (Vyse, Morley et al. 1996). In a small proportion of individuals, autoimmune diseases and/or glomerulonephritis have been reported (Vyse, Spath et al. 1994; Sadallah, Gudat et al. 1999; Amadei, Baracho et al. 2001; Baracho, Nudelman et al. 2003; Genel, Sjoholm et al. 2005) but, where available, clinico-pathological details in these cases suggest that activation of the classical pathway (possibly related to infection-driven activation of adaptive immune responses mediated by immunoglobulin production), rather than the alternative pathway, is responsible. In this situation, the presence of C3b would be predicted to allow formation of both the C3 and C5 convertases of the classical pathway provided C4b2a has been formed. Importantly, no cases of DDD in the presence of complete CFI deficiency in humans have been reported to date (Licht and Fremeaux-Bacchi 2009). Nonetheless, heterozygosity for allelic variants of CFI have been described in association with altered risk of atypical haemolytic uraemic syndrome (Fremeaux-Bacchi, Dragon-Durey et al. 2004; Caprioli, Noris et al. 2006) and C3 glomerulonephritis (Servais, Fremeaux-Bacchi et al. 2007) adding support to the view that CFI is an important component of the system regulating AP activation in humans.

Recently, mice deficient in CFI have been generated and, in common with the CFI-deficient humans, these animals exhibit reduced circulating C3 (with all the detectable C3 circulating as C3b) without features of pathological AP activation such as dense deposit disease (Rose, Paixao-Cavalcante et al. 2008). This paradox, where removal of a *negative* regulator of AP activation does not result in increased activity downstream of the AP can be explained by the theory that it is iC3b (not C3b) which is able to deposit on surfaces and mediate inflammatory effects there. This conjecture is supported by the observation that mice deficient in both CFI and CFH develop significant renal injury only after exposure to an exogenous source of CFI

(Rose, Paixao-Cavalcante et al. 2008). This presumably results in the conversion of (the plentiful) C3b into iC3b which is able to bind to the glomerular basement membrane (left unshielded by the absence of CFH), stimulate CRs which results in leukocyte recruitment and mesangiocapillary glomerulonephritis.

3.10.6 Complement Factor H

Complement Factor H (CFH) is a glycoprotein of 155 kDa which is member of the RCA cluster, is composed of 20 SCR domains and which circulates at concentrations of approximately 400 μg/ml. In the fluid phase, it is able to bind to C3b (blocking the formation of the C3 convertase, C3bBb) (Conrad, Carlo et al. 1978), has decay accelerating activity (increasing the degradation of the C3 convertase) (Weiler, Daha et al. 1976) and can act as a cofactor for Complement Factor I-mediated cleavage of C3b to form iC3b (Pangburn, Schreiber et al. 1977). In addition, CFH has the ability to bind to polyanions such as glycosaminoglycans, heparin and sialic acid molecules and this allows it to target host (as opposed to foreign) biological surfaces. Removal of sialic acid molecules from host surfaces decreases the interaction between bound C3b and CFH without impairing the interaction of C3b with Factor B or properdin, favouring C3 activation rather than degradation (Kazatchkine, Fearon et al. 1979; Pangburn, Morrison et al. 1980).

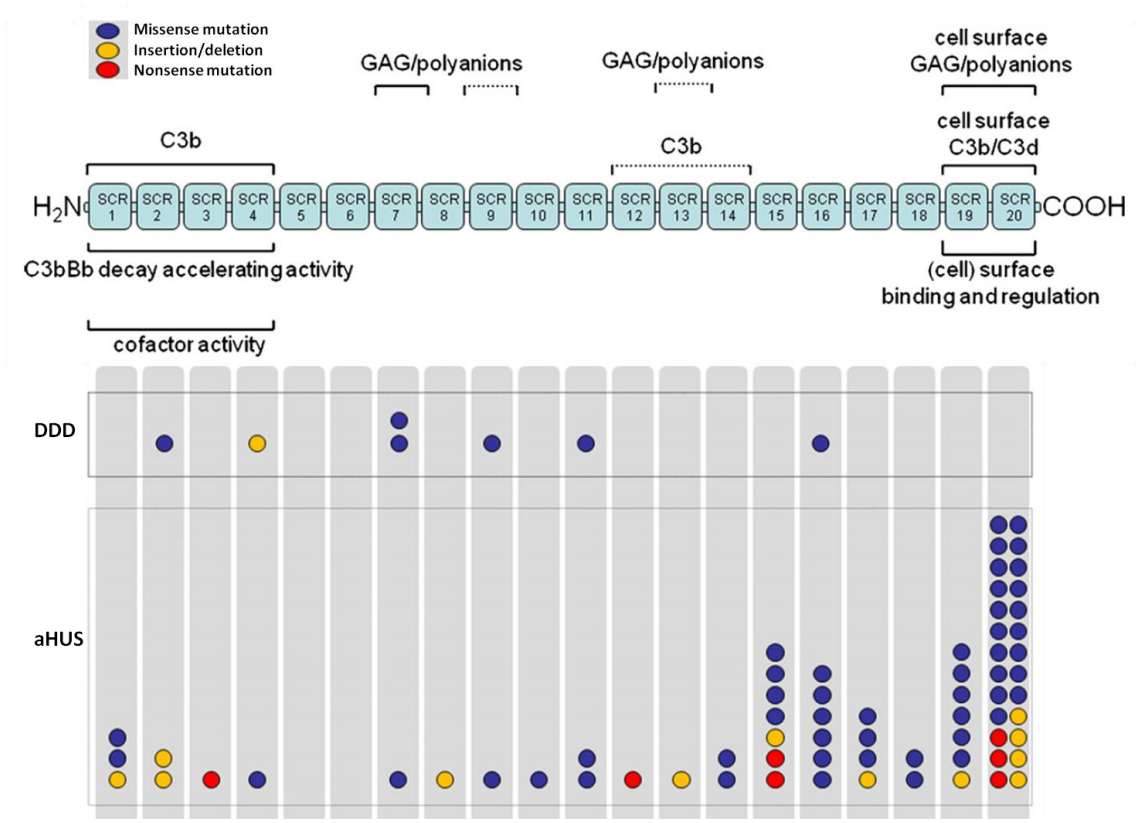


Figure 3.10-1 Complement Factor H domains and mutations associated with disease
DDD, dense deposit disease, aHUS, atypical haemolytic uraemic syndrome. Adapted from (Boon, van de Kar et al. 2009).

The functions of CFH are each conferred by properties of one or more specific regions of the protein (***Figure 3.10-1***) and their physiological importance are demonstrated firstly by the pathological consequences of mutations in each region of the gene; secondly by *in vitro* deletion-mutation biochemical studies and thirdly by transgenic murine models (Pickering and Cook 2008). These studies have indicated that the C-terminal SCRs 19 and 20 confer firstly the ability of CFH to bind to polyanions on host surfaces (Meri and Pangburn 1990) and secondly contain C3b and C3d binding sites, suggesting that this SCR confers the ability to recognise C3b in the context of self-surface. An independent C3b binding site also exists within the N-terminal SCRs 1-4 and, in addition, SCR 7 confers significant heparin-binding ability (Schmidt, Herbert et al. 2008). A construct composed only of SCRs 19 and 20 has a dominant negative effect, abolishing the ability of full length CFH to bind to surface-bound (but not fluid phase) C3b

(Ferreira, Herbert et al. 2006). Heterozygous or homozygous mutations in CFH can produce clinically significant aHUS and the observation that these mutations cluster at the C-terminal end of the protein (***Figure 3.10-1***; reviewed in (Boon, van de Kar et al. 2009)), suggests that defective complement regulation on host surfaces is an important component of this disease.

This hypothesis was supported firstly by studies of CFH in which pathogenic amino acid substitutions (associated with clinical aHUS) were introduced in SCR 20. In an elegant series of experiments the ability of these mutant proteins to protect sheep erythrocytes from complement-mediated lysis was impaired but cofactor activity in the fluid phase was indistinguishable from that of the wild type CFH protein (Sanchez-Corral, Perez-Caballero et al. 2002; Sanchez-Corral, Gonzalez-Rubio et al. 2004). A second line of evidence, using mutant mouse strains, also supported this paradigm: mice completely deficient in CFH exhibit plasma C3 consumption and C3 deposition in the glomerular basement membrane – features highly reminiscent of the human dense deposit disease (DDD) (Pickering, Cook et al. 2002). When these mice were used to generate transgenic animals in which the C-terminal 5 SCRs of CFH are deleted (ensuring that all the CFH present was devoid of SCRs 16-20), there was normal plasma C3 regulation but impaired protection of host surfaces, leading to aHUS (Pickering, de Jorge et al. 2007). In common with the murine model, complete deficiency of functional CFH in humans has been described associated with plasma C3 depletion and DDD and is not associated with any alteration in CFI levels (reviewed in (Pickering and Cook 2008; Boon, van de Kar et al. 2009)), whereas those CFH mutations causing aHUS rarely result in hypocomplementaemia (Rodriguez de Cordoba, Esparza-Gordillo et al. 2004). Notably, complete CFI deficiency is associated with reduced circulating CFH levels, and this is thought to result from CFH binding to C3b on surfaces (Pickering and Cook 2008).

Recently, a third histological category of renal disease, termed C3 glomerulonephritis (C3GN, defined by the presence of immunostaining for C3 (but not immunoglobulins) in the

glomerular mesangial areas or capillary walls in the absence of linear dense transformation of the basement membrane) has been described in association with genetic variants of CFH or CFI (Servais, Fremeaux-Bacchi et al. 2007). Whether C3GN represents an extreme (presumably mild) end of the spectrum of DDD or is genuinely a discrete clinicopathological entity remains to be determined.

3.10.7 Polymorphic variation in CFH and disease

Allelic variants of CFH have been shown to be associated with a variety of diseases, including DDD (Abrera-Abeleda, Nishimura et al. 2006) and age-related macular degeneration (AMD) (Hageman, Anderson et al. 2005). Genetic evidence also suggests a heritability of 62% for circulating CFH levels (which vary markedly in the population) (Esparza-Gordillo, Soria et al. 2004) although a relationship between circulating levels (excluding those in null mutants) and disease susceptibility has yet to be convincingly demonstrated. Recently, haplotypes within CFH associated with altered risk of DDD, AMD and aHUS have been identified and initial evidence suggests that those variants associated with increased risk of DDD are also associated with increased risk of AMD, whereas aHUS risk may be increased in the presence of variants which are not associated with increased risk of AMD/DDD (Pickering, de Jorge et al. 2007). However, linkage disequilibrium (LD) across the complex and repetitive RCA is not easily defined and it may be that the CFH variants in question are in LD with variants of nearby genes (see below) which have much greater functional influence on disease susceptibility.

3.10.8 CFH-Related Proteins

CFH lies at the 5’ end of a 360 kbp repetitive genomic region which was found to be linked with the disease in the family under investigation. This region also includes 5 additional genes known as the CFH-Related genes 1 to 5 (*CFHR1-5*) and a mutation in one of these (*CFHR5*) was found to cosegregate with the disease in the family. The CFHR proteins were originally identified by Western blotting of human serum with polyclonal antibodies raised against CFH

(Zipfel and Skerka 1994). Unsurprisingly (in view of how they were identified), all 6 genes are homologous, sharing significant sequence similarity, each being composed of multiple SCR domains. It is thought that they arose as a consequence of gene duplication events, originally of *CFH*, and it is possible to map many of the SCRs of each of the CFHRs to the homologous SCRs of CFH itself (***Figure 3.10-2***). Whether the *CFHRs* are expressed pseudogenes (with no non-redundant functional properties) or represent integral components of the sophisticated system regulating complement activity in humans has remained an open question since their identification 15 years ago, but recently several lines of evidence which may help to resolve this issue have begun to emerge.

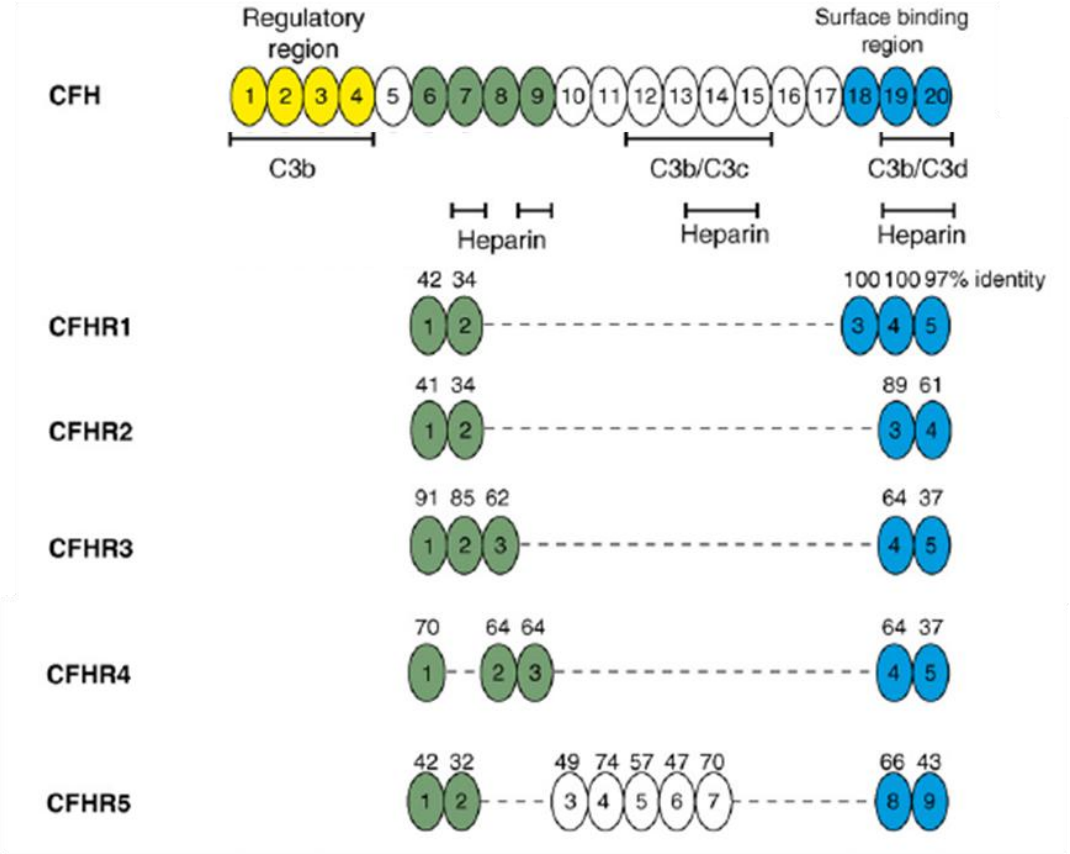


Figure 3.10-2 Homology and function of domains of CFH/CFH-Related proteins
Adapted from (Jozsi and Zipfel 2008).

CFHR1 is able to bind to the surface of pathogens (including *Borrelia burgdorferi*, *Pseudomonas aeruginosa* and Group A *Streptococci*) and to C3b, suggesting a significant complement regulating activity despite the absence of detectable cofactor activity (Skerka and Zipfel 2008). Recently, it has been shown that it can inhibit C5 convertase activity and hence the formation of the terminal complement complex (Heinen, Hartmann et al. 2009).

However, while these *in vitro* findings do suggest that CFHR1 is able to regulate complement, persuasive genetic evidence exists showing that SCRs of CFHR1, despite extensive sequence identity, have functionally and clinically apparent reduced ability to regulate AP activity when compared with the homologous SCRs of CFH: Venables and co-workers identified a family in which aHUS was caused by heterozygosity for a hybrid gene in which SCRs 19 and 20 of CFH were replaced by (the homologous) SCRs 5 and 6 of CFHR1 (Venables, Strain et al. 2006). This rearrangement produced a protein identical to CFH with the exception of 2 amino acids, and functional studies demonstrated that these substitutions resulted in a CFH protein with reduced ability to regulate complement on surfaces, and exhibited impaired binding to C3b and heparin (Heinen, Sanchez-Corral et al. 2006). It follows that these SCRs of CFHR1 are *not* required to perform the same function as the homologous SCRs of CFH.

Deletion of *CFHR1* is a common polymorphism which is associated with a reduced risk of AMD but (in the homozygous state) an increased risk of the development of autoantibodies against CFH which, in turn, are associated with increased risk of aHUS (Abarrategui-Garrido, Martinez-Barricarte et al. 2009). The mechanism by which absence of one gene can predispose to the generation of autoantibodies against the product of a homologous gene has not been conclusively established – possibilities include the generation of a neo-epitope of CFH which is present in the native form of CFHR1, or alternatively there may be an immunogenic CFH allele which is in LD with the CFHR1 deletion. The second explanation seems less likely since additional CFHR1 deletions and truncation mutations (i.e. independent alleles) have been

identified in association with the development of CFH antibodies (Abarrategui-Garrido, Martinez-Barricarte et al. 2009).

CFHR5, which is the largest of the CFHR proteins and comprises 9 SCRs, has been identified within complement deposits in the rat glomerulus (McRae, Cowan et al. 2001). It has been shown to bind C3b and heparin and to have cofactor activity at concentrations of several hundred ng/µL (McRae, Duthy et al. 2005). Common variants of CFHR5 are preferentially represented in cohorts of patients with DDD (Abrera-Abeleda, Nishimura et al. 2006) and aHUS (Monteferrante, Brioschi et al. 2007), consistent with an important and non-redundant role for the protein, but in the absence of conclusive functional or genetic data it remained unclear whether these observations result from LD leading to the detection of effects of risk variants elsewhere within the RCA. The work presented here, which indicates that a mutation in *CFHR5* is responsible for a highly penetrant Mendelian disease, provides direct evidence that CFHR5 has a non-redundant role in the regulation of complement in humans.

3.11 A monogenic cause of C3 Glomerulonephritis

3.11.1 Family 1

The index case (designated V-4 in ***Figure 3.11-1***) was born in the UK to Cypriot parents. He presented to his GP at the age of 9 with a two year history of headaches. He had had a right orchidopexy in infancy and there was no other significant medical or surgical history. The GP noted that he had not had urinary tract infections in the past. At the time of this review he was noted to be "rather thin and frail" and was found to be pyrexial with an upper respiratory tract infection (URTI). The blood pressure was 140/90 and he was referred for a paediatric opinion, for which there are no available records.

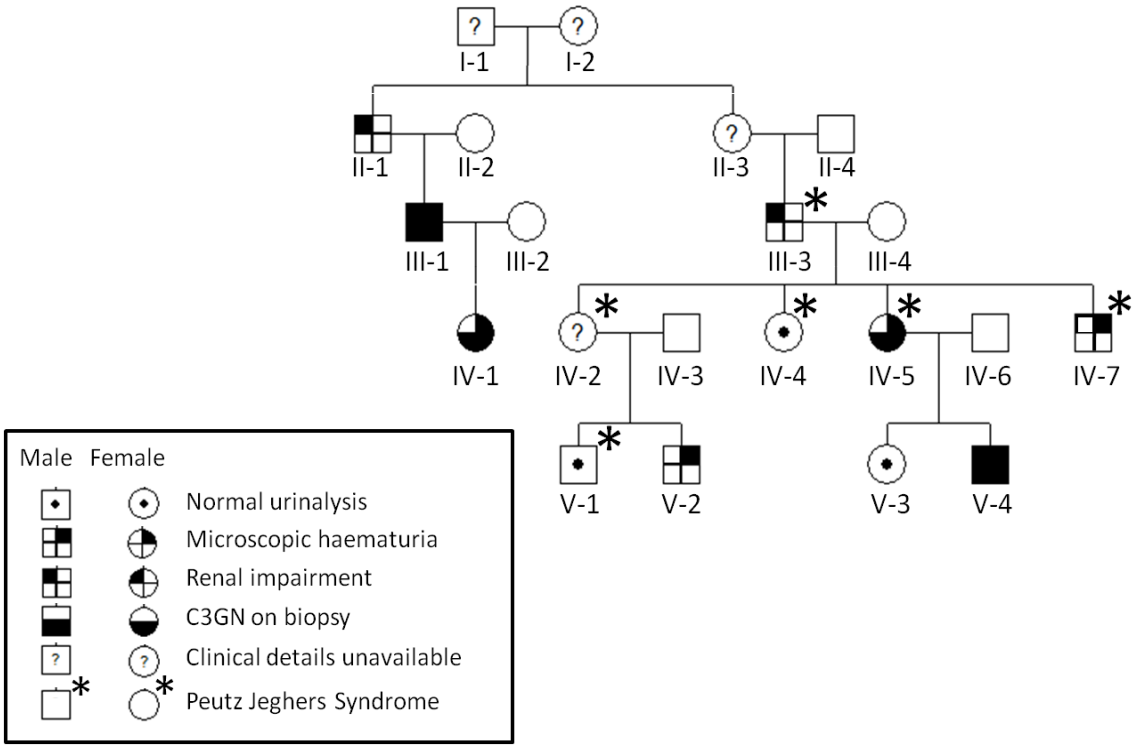


Figure 3.11-1 Family 1 pedigree

He remained well (with no headaches) until the age of 14 when he was seen by a cardiologist for investigation of palpitations. Urinalysis at this time demonstrated a trace of blood but no other abnormalities were found and he was reassured. He was referred to the

nephrology service at St Mary's Hospital in London two years later because of hypertension. At this time he was symptomatically "very well". Physical examination was entirely normal with the exception of an elevated blood pressure of 165/80 mmHg and urinalysis, which demonstrated +++ blood only. Circulating biochemical and haematological indices were all within the normal ranges. Serum levels of immunoglobulin (Ig) A, IgG and IgM were normal and complement C3 was 1.05 g/l (NR 0.73-1.77). An autoantibody screen, including antinuclear and anti-neutrophil cytoplasmic antibodies and C3 nephritic factor, was negative and ^{51}Cr-EDTA clearance was 100 ml/min.

A clinical diagnosis of "likely IgA nephropathy" was documented by his nephrologist at this time and he underwent a renal biopsy. Light microscopic examination revealed two glomeruli with tuft collapse and thickening of Bowman's capsule and 10% focal atrophy in the areas adjacent to these glomeruli. Immunoperoxidase staining revealed C3 deposition on the capillary wall and in a few glomeruli. There was no staining with antibodies for IgA, IgM, IgG or C1q. Electron microscopy revealed expansion of the mesangial matrix with some peripheral electron dense deposits. There were multiple elongated subendothelial dense deposits in the glomerular basement membrane but no subepithelial deposits were observed.

The blood pressure was controlled with enalapril therapy and he remained well until the age of 20 at which time he reported sudden onset of abdominal pain, with dark stools noticed one day later. The haemoglobin was 7.8 g/dl and he was admitted to his local hospital. An oesophago-gastro-duodenal endoscopic examination revealed no source of bleeding and no polyps. Serum creatinine at this time was 89 µmol/l and there were no clinical features to suggest Peutz Jeghers syndrome. There have been no further episodes of gastrointestinal bleeding.

At the age of 26 he was found to have proteinuria ++ in addition to haematuria and the glomerular filtration rate was 60 ml/min. Urinary protein loss was less than 1 g per day. A

second renal biopsy was performed: light microscopic examination revealed increased mesangial cells and matrix with segmental capillary wall thickening. Immunostaining revealed course granular C3 deposition on the capillary wall and electron microscopy again demonstrated elongated subendothelial electron dense deposits with occasional subepithelial deposits (***Figure 3.11-2***). At the time of writing he was well on antihypertensive medication aged 31.

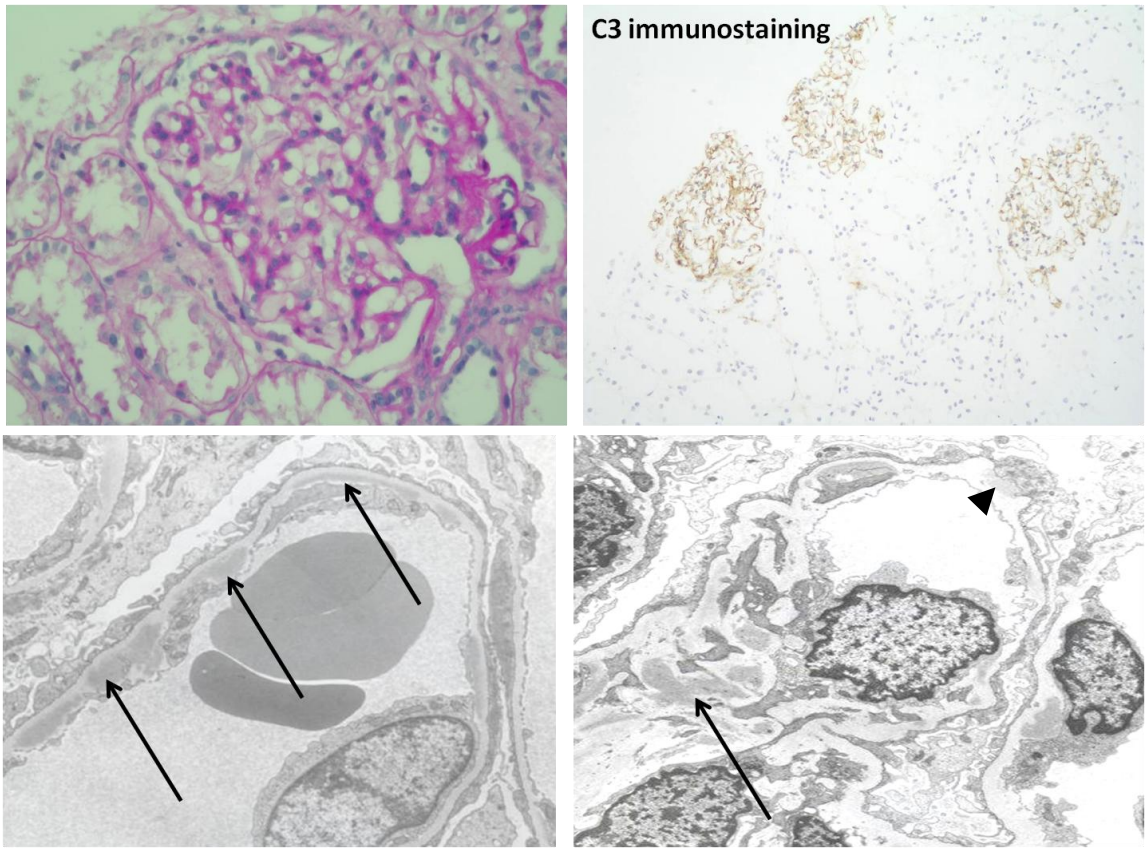


Figure 3.11-2 Kidney biopsy in individual V-4
Arrows show electron dense deposits in the subendothelial and mesangial regions. A single subepithelial hump is visible (arrowhead).

IV-5 (the mother of V-4, ***Figure 3.11-1***) was born in the village of Gerakies in the Troodos region of Cyprus. She had exhibited perioral and buccal pigmented macules in youth which had faded with age, and a small number of irregular, splinter shaped pigmented macules on the fingers (***Figure 3.11-3***). Multiple gastrointestinal polyps had been removed endoscopically over

the course of her adult life and histological examination of the polyps had revealed cystic-glandular hyperplasia, consistent with Peutz Jeghers syndrome (***Figure 3.11-4***). Bilateral oopherectomies were performed at the age of 51 and histological examination revealed a left serous cystadenoma. Her father (III-3) had been diagnosed with Peutz Jeghers syndrome prior to his death from ischaemic heart disease.

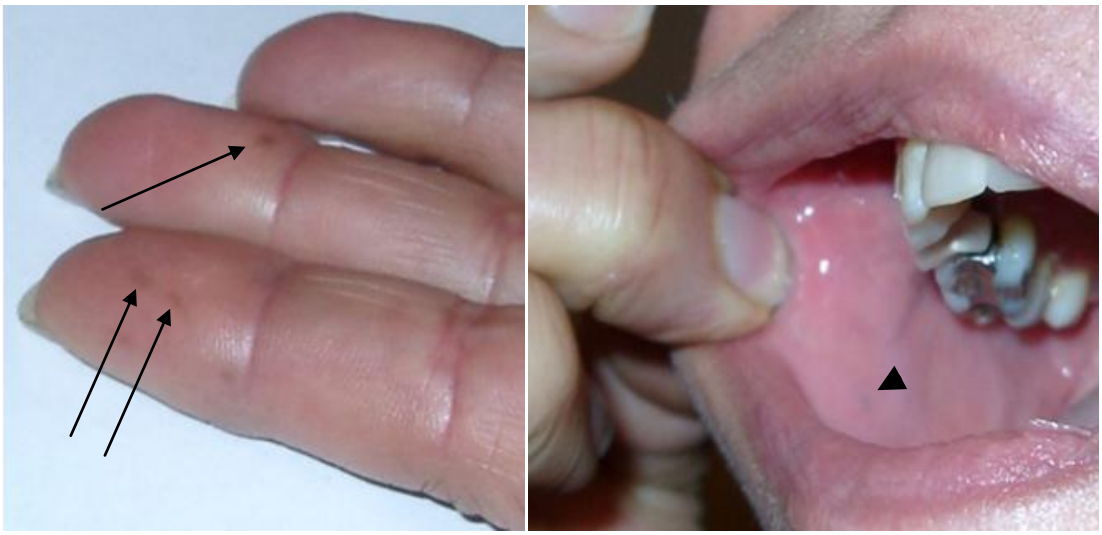


Figure 3.11-3 Clinical features of Peutz Jeghers syndrome
Pigmented papules visible on the fingers of individual IV-5 (arrows). A faint papule was also visible on the buccal mucosa (arrowhead).

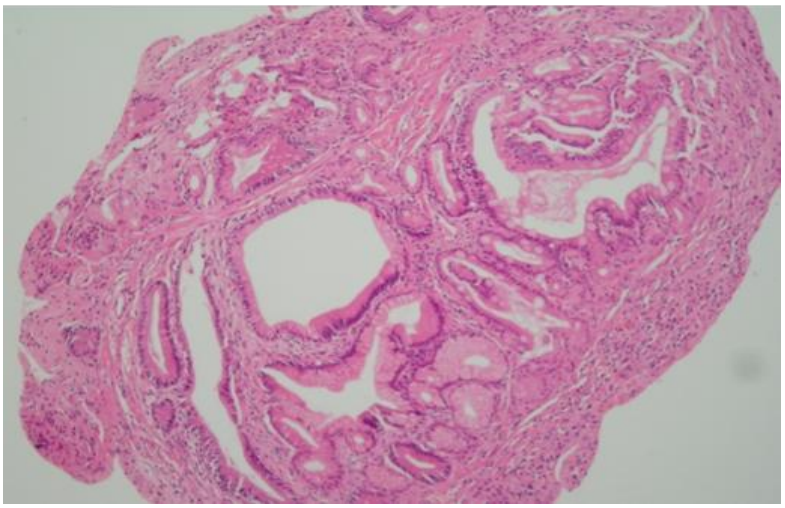


Figure 3.11-4 Histological features of Peutz Jeghers Syndrome
Cystic glandular hyperplasia in a histological section from a gastric polyp removed from individual IV-5.

IV-5 had been referred to the nephrology service at St Mary's Hospital, London at the age of 40 with persistent microscopic haematuria. Biochemical and haematological parameters were all within the normal ranges with a serum creatinine of 90 µmol/l. Circulating complement C3 level was 1.08 g/l. Neither autoantibodies nor C3 nephritic factor were detected in the serum and immunoglobulin levels were normal. She underwent renal biopsy which demonstrated slight mesangial expansion with occasional capillary loop occlusion. There was 5% tubular atrophy. Electron microscopy demonstrated scattered, small subendothelial electron dense deposits and a few larger subepithelial deposits. There were occasional mesangial electron dense deposits with mesangial cell interposition.

A paternal relative (III-1 in ***Figure 3.11-1***), who had emigrated from Gerakies to London some years previously with his parents had presented to the nephrology service of Charing Cross Hospital with haematuria and proteinuria. A renal biopsy was performed at the age of 43 which showed mesangiocapillary glomerulonephritis (MCGN) with immunostaining for C3 but not immunoglobulins or C1q. He commenced haemodialysis at the age of 46 and received a cadaveric renal allograft one year later. He died 10 years subsequent to this from a myocardial infarction at which time there was good allograft function – the graft was never biopsied and serum creatinine prior to death was 89 µmol/l. There was a history of haemorrhagic gastritis but no features of Peutz Jeghers syndrome were documented. Of note, his father (II-1) had developed end-stage renal disease and was receiving haemodialysis prior to his death, however further clinical details are not available.

3.11.2 Independently identified patient from Cyprus

An unrelated individual was referred from Nicosia to Hammersmith Hospital for a second opinion. At the age of 31 he had been found to have microscopic haematuria, with hypertension first noted at age 35, at which time his serum creatinine was 103 µmol/l. Over the following 4 years he reported recurrent episodes of synpharyngitic macroscopic haematuria,

one of which directly preceded a sharp rise in his serum creatinine to 237 µmol/l, recovering to 172 µmol/l some months later. Urine protein loss was relatively constant at around 400 mg/day. Tests for circulating antibodies including antinuclear and anti-neutrophil cytoplasmic antibodies and C3 nephritic factor, were negative. At age 38 a renal biopsy was performed. Light microscopic examination revealed expansion of mesangial cells and matrix and 10% focal tubular atrophy. Immunostaining was positive for complement C3 and C9 but negative for immunoglobulins. Electron microscopy revealed mesangial, subendothelial and a few scattered subepithelial electron dense deposits. The histopathologist (Professor Terry Cook) considered the biopsy findings highly unusual but noticed the similarity with those seen in individuals from family 1. No features of Peutz Jeghers syndrome were observed and a family history of renal disease was not documented.

In summary, I have described a multiply affected kindred with ancestry in the Troodos Mountains of Cyprus in which there is autosomal dominant inheritance of renal disease characterised by glomerular C3 deposition in the absence of significant immunoglobulin or C1q – appearances consistent with C3 glomerulonephritis (C3GN). There are characteristic C3-containing subendothelial, mesangial and occasional subepithelial deposits visible on electron microscopy and all affected individuals have microscopic haematuria. Sub-nephrotic range proteinuria and renal impairment are observed in some individuals. The clinico-pathological features were strikingly reminiscent of IgA nephropathy with the exception of a) autosomal dominant inheritance and b) a lack of IgA staining on renal biopsies in all affected individuals. Peutz Jeghers syndrome is also present in the family although it did not appear to cosegregate with the renal disease. In addition, an unrelated individual from Cyprus was also found to have C3GN, with renal biopsy findings remarkably similar to those seen in family 1. In addition to microscopic haematuria, hypertension and renal impairment, he reported repeated episodes of synpharyngitic macroscopic haematuria which appeared to coincide with stepwise deterioration of kidney function.

3.12 Investigation strategy

The aim of this investigation was to identify the gene responsible for the kidney disorder affecting this family. While the phenotype described is not of a well-recognised Mendelian disorder, the deposition of C3 in the glomerulus (in the absence of immunoglobulins and C1q) that was present in all 4 affected individuals who were biopsied suggested that a defect of complement alternative pathway regulation was present. Genes implicated in other disorders of complement regulation which result in kidney disease include Complement Factor H (CFH), Factor B, Complement Factor I (CFI), Complement C3, Membrane Cofactor Protein (MCP). The exons and flanking intronic sequences of these genes were sequenced in IV-5 and V-4 from family 1 prior to my involvement in this investigation by colleagues of Dr Pickering, and no likely pathogenic variants were identified.

Review of the available clinicopathological information highlighted the following salient points:

1. Glomerular deposition of C3 in the absence of immunoglobulins in the observed pattern (i.e. C3GN) is exceedingly rare, and coincidental occurrence of these biopsy appearances in multiple members of a single family, in the absence of a shared genetic or environmental aetiological agent, was unlikely enough to be discounted. The possibility that the independently identified individual from Cyprus was a relative of family 1 was intriguing and would potentially strengthen a linkage study.
2. V-4 (family 1) was born and brought up in the UK so I considered an environmental factor restricted to the Troodos Mountains unlikely to explain the observations.
3. The independently identified individual from Cyprus reported synpharyngitic macroscopic haematuria – a symptom considered almost pathognomonic of IgA nephropathy. Furthermore, histological features were compatible with IgA

nephropathy in all cases, with the obvious exception of the absence of detectable IgA deposition.

4. Peutz Jeghers syndrome (PJS) is a rare condition (affecting between 1 in 60,000 and 1 in 300,000 of the population). Coincidental occurrence of 2 unrelated rare disease, while not impossible, appeared unlikely. Notably, PJS was not present in III-1 and V-4 from family 1, nor in the independently identified patient from Cyprus.

In order to answer these questions, I proceeded with the following steps:

1. Interview affected individuals to identify the presence of:
    a. synpharyngitic macroscopic haematuria
    b. additional relatives with renal disease and/or PJS
2. Screen relatives of affected individuals for the presence of digital, perioral or buccal pigmentation; a history of gastrointestinal bleeding or intussusceptions; urinary abnormalities (haematuria and proteinuria); or renal impairment
3. Perform retinal photography in affected individuals to look for extra-renal manifestations which might be expected in the presence of a defect of AP regulation
4. Candidate gene sequencing:
    a. Amplify and sequence exons of *STK11* in IV-5 and V-4 from family 1
    b. Amplify and sequence exons of *CFHR5* in IV-5 and V-4 from family 1
5. Collect DNA and serum from consenting family members, whether affected, unaffected and unknown affection status, genotype 6000 SNPs across the genome and perform a linkage study in order to inform the selection of additional candidate genes to investigate

3.13 Family screening and DNA collection

3.13.1 Family 1

On direct questioning, individual V-4 reported repeated episodes of brown-red discolouration of the urine which occurred immediately preceding or within 1-2 days of the onset of upper respiratory tract infections. These episodes of synpharyngitic macroscopic haematuria usually lasted for between 2 and 5 days and had occurred approximately once or twice each year since the age of 14. Individual IV-5 could recall having experienced one or two similar episodes over the course of her life.

IV-5 from family 1 reported that her father's cousin's daughter (IV-1 in ***Figure 3.11-1***) also had kidney disease. Individual IV-1 had been referred to Charing Cross Hospital 20 years previously at the age of 25 because of persistent microscopic haematuria with a trace of proteinuria. Renal biopsy at that time had demonstrated focal interstitial fibrosis and some collapse of glomerular capillary loops. Immunostaining demonstrated the presence of diffuse, intense granular deposition of C3 along the capillary loops with scanty staining for IgM and IgA but not IgG or C1q. Electron microscopy revealed conspicuous, elongated subendothelial electron dense deposits with a few discrete subepithelial deposits. Deposits were also present in the mesangial regions which were otherwise normal. Immuno-gold electron microscopy demonstrated that the electron dense deposits were bound by antibodies directed against complement C3 (***Figure 3.13-1***). She had remained entirely well during this period, with no hypertension, renal impairment or episodes of macroscopic haematuria reported. She had no clinical stigmata of Peutz Jeghers Syndrome.

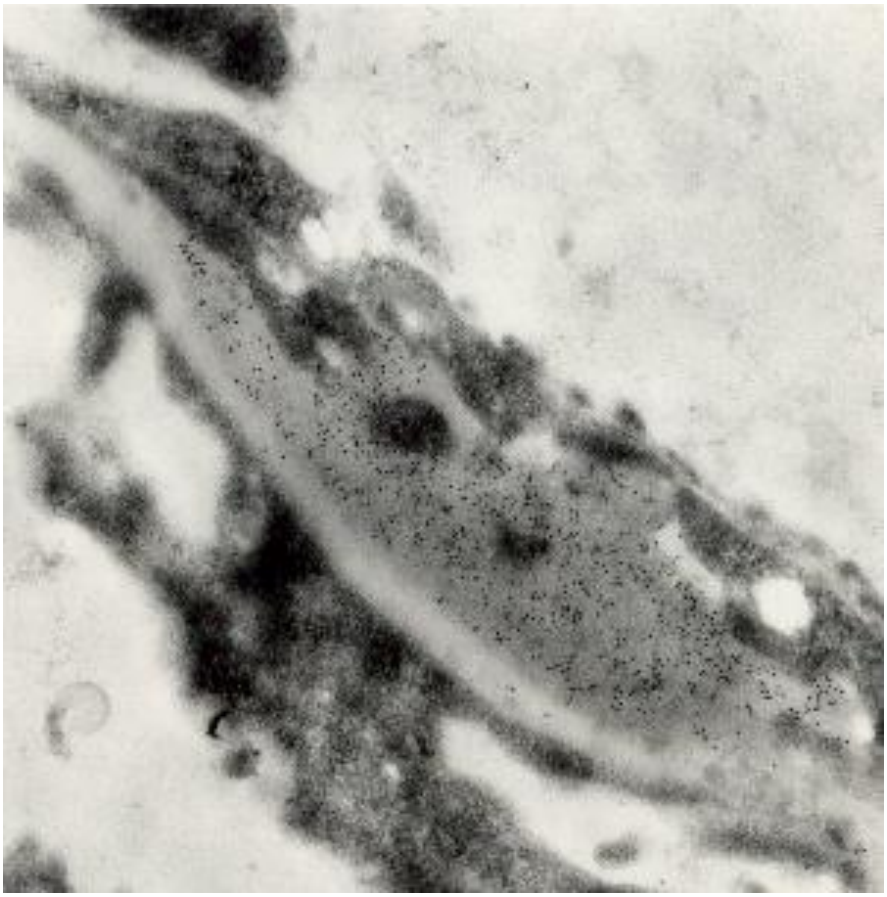


Figure 3.13-1 Immuno-gold electron microscopy
Immuno-gold electron micrograph of a kidney biopsy from individual IV-1. Black dots indicate immunoreactivity for Complement C3 in a subendothelial electron dense deposit

At the invitation of IV-5, I travelled to Cyprus to meet additional relatives who reside there. Urinary dipstix testing identified + haematuria and trace proteinuria in individual IV-7. He also exhibited perioral pigmentation and had been diagnosed with Peutz Jeghers syndrome. Individual IV-2 (sister of IV-5, ***Figure 3.11-1***) had died from disseminated intra-abdominal malignancy in her early forties. One of her sons (V-1) had been diagnosed with PJS owing to the presence of digital, buccal and perioral pigmentation (***Figure 3.13-2***) and numerous gastrointestinal hamartomas which had been removed endoscopically. Urinalysis was normal. His brother (V-2) exhibited no stigmata of PJS and was entirely well. Urinalysis revealed a trace of blood only. IV-4 (also sister to IV-5, ***Figure 3.11-1***) had subtle perioral pigmentation

gastrointestinal polyps and had been diagnosed with PJS some years previously. Urinalysis was normal. None of these individuals reported any episodes of macroscopic haematuria.

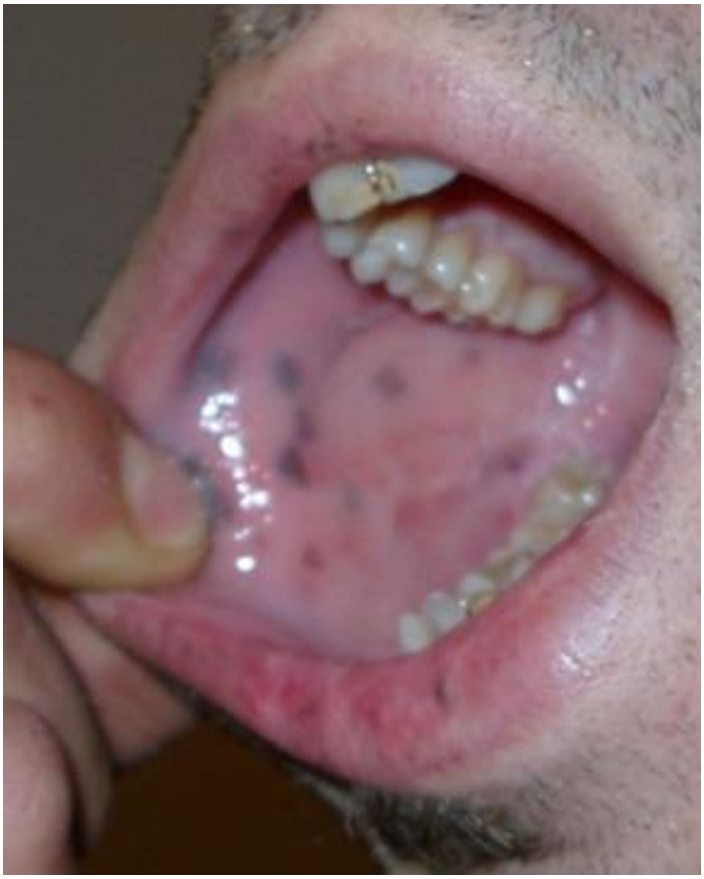


Figure 3.13-2 Pigmented macules on the lips and buccal mucosa of individual V-1

3.13.2 Family 2

On direct questioning, the unrelated individual from Cyprus (III-2 in ***Figure 3.13-3***) reported that, while his father was from Larnaca in Eastern Cyprus, his mother's family was from Kalopanagiotis, which lies approximately 3km from, and in the same valley as, Gerakies in the Troodos Mountains (***Figure 3.13-4***). Her father (I-1) had died from kidney failure at the age of 65 in 1953 and her brother (II-3) had recently died with many medical problems, including 'severe' kidney disease in his 60s. Further details from these individuals were not available. III-2 invited me to Cyprus where I was able to meet and screen additional relatives. This revealed microscopic haematuria in his mother, aunt and 2 cousins (individuals II-1, II-4, III-4 and III-5).

All these individuals had recent serum creatinine measurements which were within the normal range and no evidence of proteinuria. None had been biopsied. In addition, the daughter of one of the cousins (IV-1) was subsequently found to have microscopic haematuria. None of these female relatives could recall any episodes of macroscopic haematuria. Furthermore, no individuals from family 2 exhibited abnormal pigmented macules and there was no history of GI bleeding, polyps or intussusception.

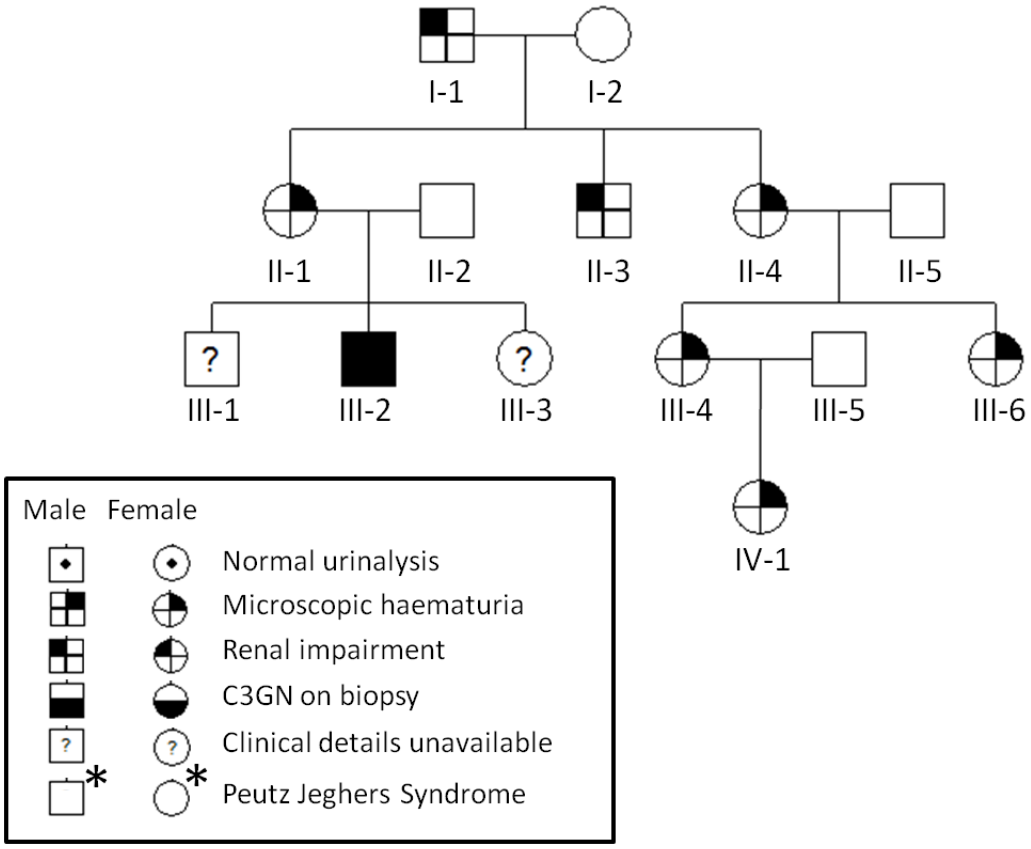


Figure 3.13-3 Family 2 pedigree

None of the known surnames of ancestors of family 2 were shared by known ancestors of family 1 and neither family could trace ancestry to the other village. Several members of the family pointed out to me that marriages between the residents of the two villages were extremely common. In addition, individual III-4 from Family 2, informed me that the Kikkos Monastery was situated near the villages and that it was common folklore that many of the

residents of the valley were the product of (illegitimate) liaisons between monks and villagers. I could find no evidence to lend direct support to this speculation and therefore considered that, if the two families did have an affected common ancestor, this individual had probably lived before living memory.

All individuals from both families were eager to participate in the research.

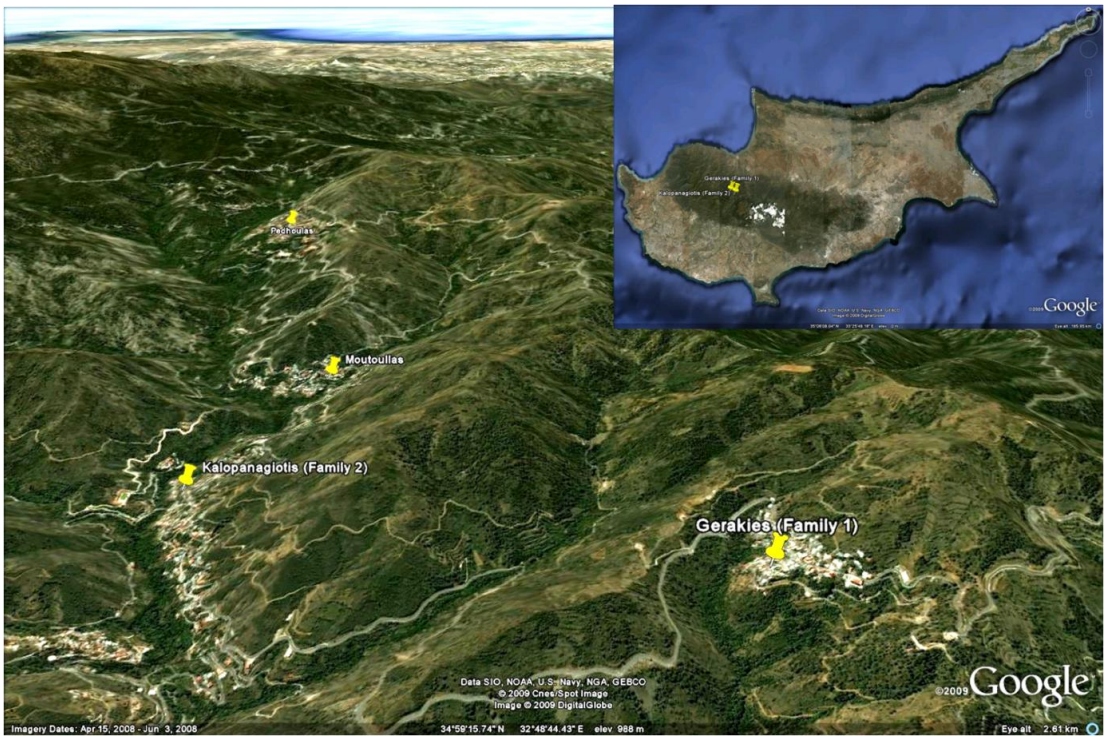


Figure 3.13-4 Satellite photograph of the Troodos region

Showing the region from which the families reported ancestry, and their location on the island of Cyprus (inset). Images downloaded from www.google.com.

3.14 Retinal photography

Since defects of complement regulation have been associated with retinal abnormalities (particularly drusen in dense deposit disease (Duvall-Young, MacDonald et al. 1989)), retinal photography was performed in IV-5 and V-4. This demonstrated no abnormality (***Figure 3.14-1***).

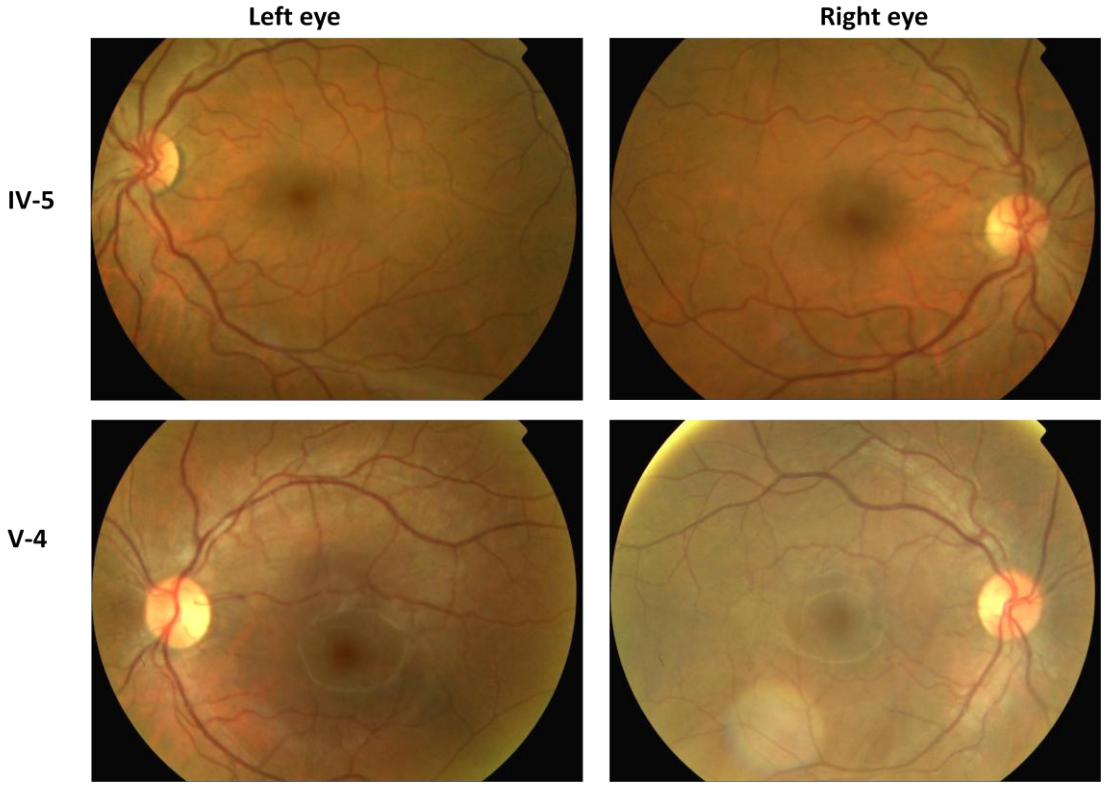


Figure 3.14-1 Normal retinal appearance of IV-5 and V-4 (family 1)

3.15 Candidate gene sequencing:

3.15.1 STK11

Peutz Jeghers syndrome (PJS, OMIM #175200) is a rare, autosomal dominant disorder characterised by perioral, buccal and digital pigmented macules and multiple gastrointestinal hamartomas. It was first described in 1921 and extra-intestinal polyps recognised as a feature of the disease in the 1960s. Affected individuals are at increased risk of a large number of neoplasms, both intra- and extra- intestinal, including rare ovarian, cervical and testicular tumours and breast carcinoma. Mutations in the *Serine Threonine Kinase 11* (*STK11*) gene on chromosome 19p13.3 have been identified cosegregating with PJS in numerous families, with loss of heterozygosity (i.e. somatic inactivation of the second allele) detectable in 70% of polyps studied (Gruber, Entius et al. 1998). At the time this part of the investigation was performed (summer 2007), no families with PJS had been reported in the published literature in which a mutation in a gene other than *STK11* had been implicated in causing the disease. However, in 2008, A single individual with clinical features reminiscent of PJS was reported in whom a mutation in the *Myosin Heavy chain 11* (*MYH11)* gene on chromosome 16p13.13 was identified (Alhopuro, Phichith et al. 2008).

Intriguingly, the gene for Complement C3 lies 5 megabase-pairs (Mbp) downstream of *STK11* on the short arm of chromosome 19. It therefore seemed possible that a large inversion had occurred resulting in disruption to the two genes and causing both the renal disease and PJS, with incomplete penetrance of each phenotype. Other possibilities included:

1. The presence of an *STK11* mutation linked to a C3 mutation in family 1 with a recombination event having occurred since families 1 and 2 diverged
2. Both PJS and the renal disease were caused by a single mutation with incomplete penetrance of both diseases

3. Completely coincidental inheritance of 2 unlinked mutations in some members of family 1 only, with the PJS mutation absent in family 2.

Possibility 2 (the same mutation in an unknown gene causing both diseases) was hard to envisage because the biology of PJS (which results from a defect of intracellular energy sensing) and complement (predominantly a cascade of circulating proteins secreted by the liver) seem so diverse. Nonetheless, it would not be impossible for a trafficking, gain-of-function or gene fusion mutation to have effects on such disparate systems.

In order to disentangle these possibilities, firstly cytogenetic analysis (with particular emphasis on chromosome 19p) of lymphocytes from individual IV-5 was performed at the Kennedy-Galton centre for clinical genetics, Northwick Park Hospital. This was reported as normal. Second, IV-5 and V-4 were screened for *STK11* mutations. This was clinically important because V-4 had experienced a gastrointestinal bleed, yet had no clinical evidence of PJS and a molecular diagnosis might therefore be valuable in guiding decisions on further gastro-intestinal surveillance in him (as well as other ostensibly unaffected relatives). As a consequence, exons of *STK11* in were amplified from genomic DNA from IV-5 and V-4 and sequenced. This revealed a novel single C to A substitution at a position 1262 base pairs from the transcription start site in exon 1 of the gene (C1262A), present only in IV-5 (***Figure 3.15-1***). This predicts a change from Tyrosine 49 to a premature termination codon (Tyr49X), and is therefore highly likely to be pathogenic. V-4 only had evidence of the database allele at this position.

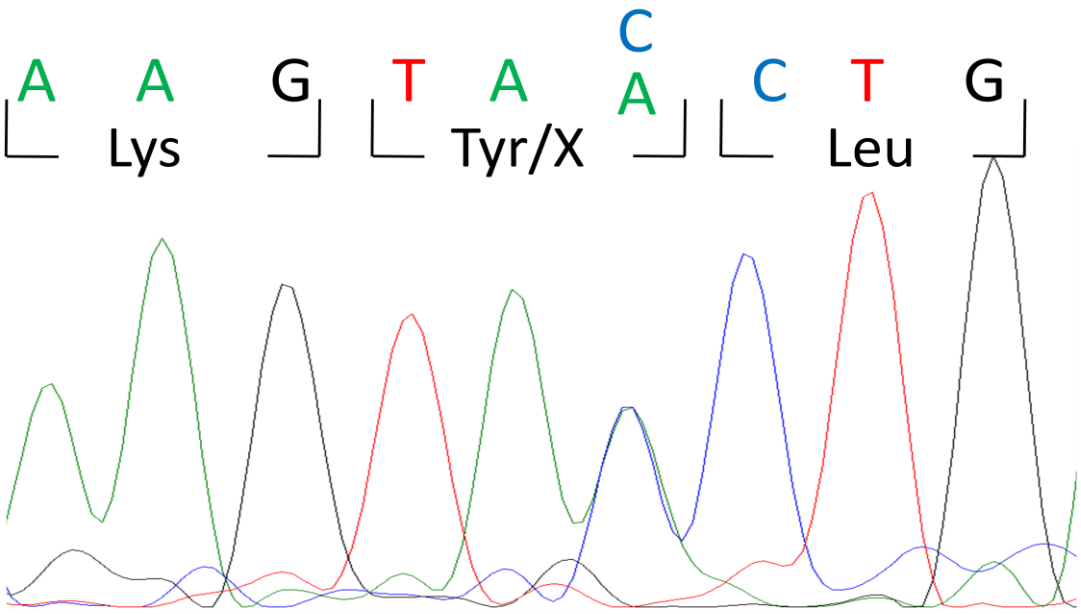


Figure 3.15-1 Premature stop codon in exon 1 of STK11
Detected in DNA from individual IV-5 from family 1.

C1262A is predicted to abolish an *AfaI* restriction site in a PCR amplicon generated by amplification of exon 1 of *STK11.* This allowed screening of clinically affected and unaffected members of the families which identified the mutation in all and only individuals with a clinical diagnosis of PJS (***Figure 3.15-2***). This allowed calculation of a logarithm of odds ratio (LOD score) of 2.1 for segregation with PJS among descendants of III-3. It was not identified in any members of family 2 or individuals IV-1 and V-4 from family 1, proving that the renal disease did not result from this *STK11* mutation.

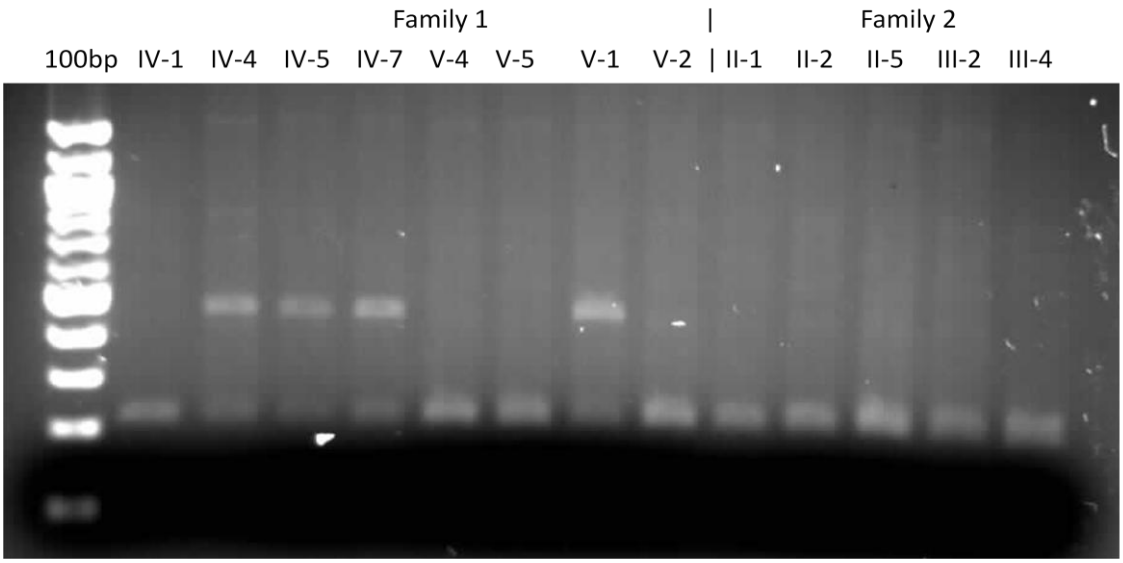


Figure 3.15-2 Screening for STK11 mutation
Restriction digest showing heterozygosity for abolished *Afa1* restriction site only in individuals with clinical evidence of PJS.

In addition, this pattern of recombinations made it unlikely that the mutation responsible for the renal disease was located near to *STK11,* since 1 or more recombinations would have had to occur between III-3 and IV-1 (5 meioses) AND there would have had to be at least 1 recombination between III-3 and V-1 (2 meioses) AND 1 recombination between III-3 and IV-4 (1 meiosis) AND 1 recombination between IV-5 and V-4 (1 meiosis) AND no recombinations between III-3, IV-5 and IV-7 (2 meioses).

The likelihood of this occurring at a given recombination fraction, θ, can be calculated as follows:

$$L(Observed|\theta) = (1-(1-\theta)^5) \times (1-(1-\theta)^2) \times \theta^2 \times (1-\theta)^2$$

It follows that:

$$LOD\ score = \log_{10} \frac{L(Observed|\theta)}{L(Observed|0.5)}$$

This can be plotted graphically (***Figure 3.15-3***)

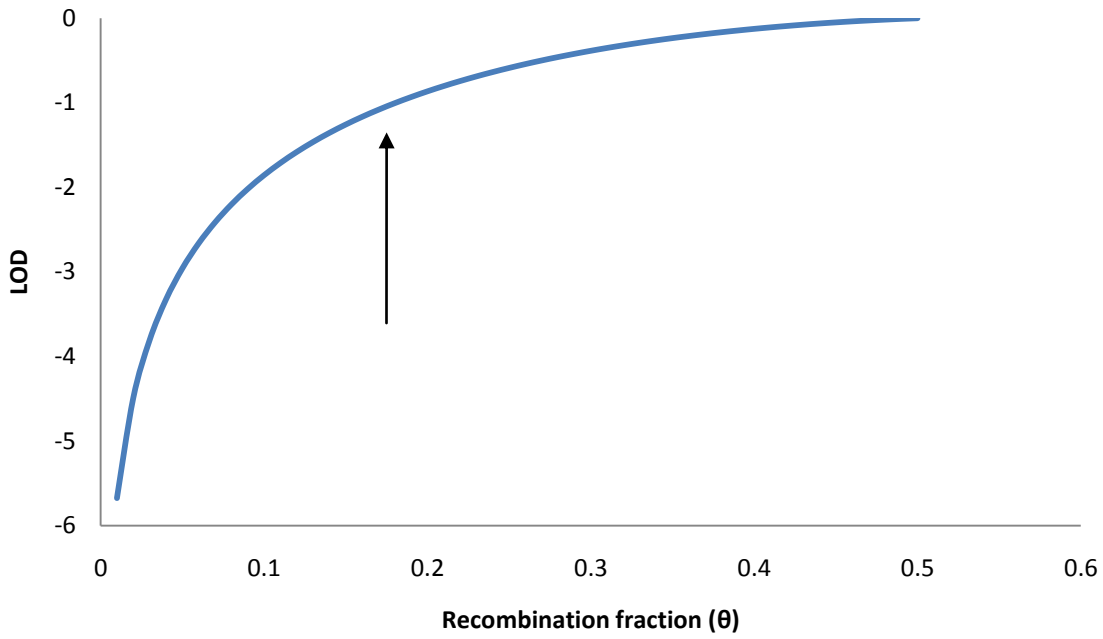


Figure 3.15-3 Linkage of renal disease with STK11
The LOD score is -0.97 at θ=0.185, the predicted position of the *C3* gene (arrow).

Since the *C3* gene is situated 5 megabases (corresponding, in this region of the genome, to 18.5 cM[a]) from *STK11* the predicted recombination fraction is approximately 0.185, which yields a LOD of -0.97, indicating that linkage between the renal disease and *C3* was unlikely.

3.15.2 CFHR5

As discussed above, the coding and flanking regions of genes previously implicated in complement alternative pathway dysregulation leading to renal disease (namely CFH, CFI, CFB, C3, MCP) had been sequenced without indentifying any likely pathological variants. Recent work (Abrera-Abeleda, Nishimura et al. 2006; Monteferrante, Brioschi et al. 2007) had implicated variation in CFH Related protein 5 in modifying the risk of dense deposit disease (DDD) and haemolytic uraemic syndrome respectively. These were both case-control studies in which certain SNP alleles were over-represented in cases compared to controls. While this sort

[a] deCODE genetic map

of study is unlikely to confirm an aetiological role of a particular variant (firstly because the effect of other alleles in linkage disequilibrium with those directly measured could mediate any biological effect and secondly because, since the variants are present in a significant proportion of healthy controls, any functional effect on protein function or expression is likely to be subtle and hence difficult to study *in vitro*), this work does provide evidence that CFHR5 *could* play a role in AP regulation and therefore I considered sequencing the gene worthwhile.

Sequencing was first attempted by amplification of exons of CFHR5 using the primers published in (Abrera-Abeleda, Nishimura et al. 2006), however this was unsuccessful and I selected new primers which did allow amplification and sequencing of the gene. This was performed in IV-1 and IV-5 from family 1 and in 2 unrelated controls. No likely pathogenic coding sequence or splice-site variants were detected and heterozygosity for a database SNP was observed in IV-1 (although IV-5 was homozygous for all). This excluded loss of one copy of the entire gene in an individual with biopsy-proven C3GN.

Further candidate genes (such as *CFHR1-4*) were not sequenced at this time as I took the view that a linkage study would be a more efficient way of targeting continued efforts. In particular, since there was no evidence of circulating complement activation/depletion and no indication of complement deposition outside the kidney, it seemed possible that an abnormality of a gene which is usually expressed in the kidney (i.e. which is outwith the complement pathway) could be responsible for localised complement activation there.

3.16 Genome-wide linkage study

Simulation analysis (using FastSLink) indicated that, with the samples available, the maximum predicted LOD score (assuming linkage would be identified at the same locus in both families) would be 3.67 (***Figure 3.16-1***). While a LOD of 3.67 would be considered significant for linkage by conventional criteria, this information was not critical in determining whether a linkage study was worthwhile. In fact, the prime rationale for performing the linkage analysis was to *exclude* genes which had been prior candidates (and had no coding/splice site mutations detected so far). This is because the two most likely options were a) a mutation in a gene not previously considered (in which case *all* the candidate genes might be excluded on the grounds of linkage) or b) a mutation in a gene previously shown to have no coding/splice site mutations in this family (i.e. some form of copy number variation too large to be captured in a single PCR amplicon). This eventuality would result in at least one locus containing a candidate gene being linked to the renal disease.

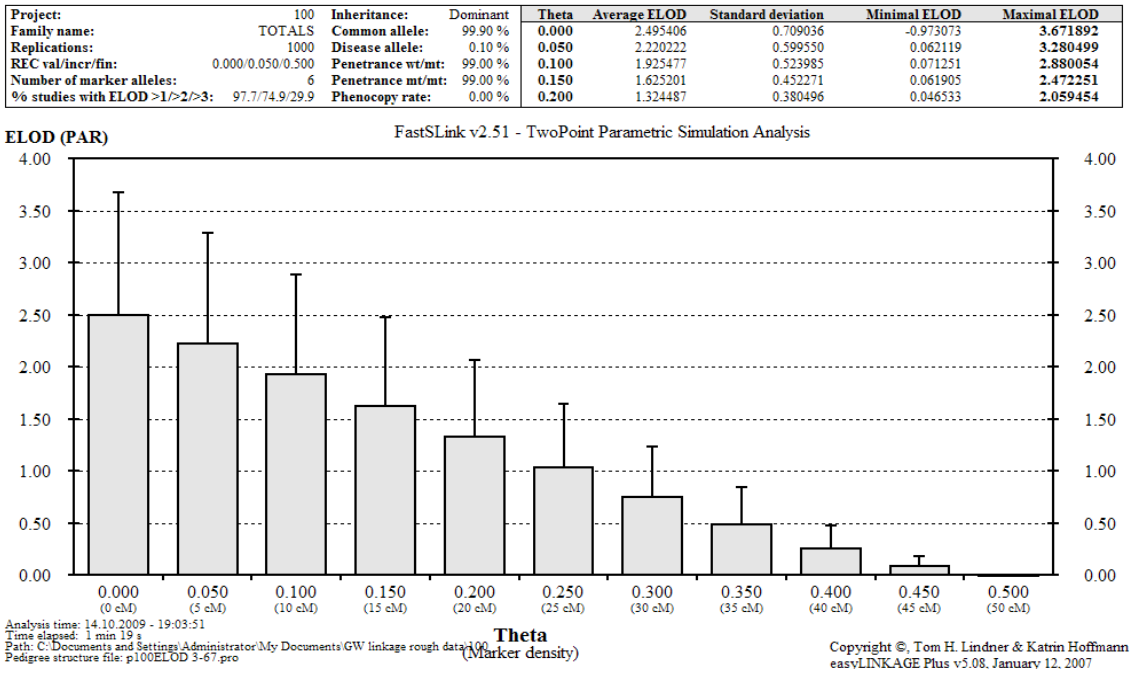


Figure 3.16-1 Simulation analysis
Simulation using FASTSLINK indicated maximum predicted LOD score of 3.67 for families 1 and 2.

SNP genotyping was performed using the Illumina Linkage IV array which comprises 6008 SNPs distributed across the genome with an average spacing of 0.32 cM. The cost-benefit analysis of using this chip was favourable at the time the experiment was performed.

Family members were scored as unaffected if they married into one of the families, or if they had a normal urinalysis and no renal impairment. Individuals were scored as affected if they had renal impairment or microscopic haematuria. Renal biopsies showing C3GN were available from 4 individuals for whom DNA was available.

46 SNPs with Mendelian errors were identified across the families. These SNPs were removed from subsequent analyses. This represents less than 0.1% of all genotypes and was consistent with accurate sample labelling, genotyping and data transfer.

Genome-wide linkage analysis excluded linkage with renal disease with a LOD of < -2 in 89.6% of the genome. This included the loci containing *CFI*, *CFB*, *MCP*, *C3*, *STK11.* Of the 10.4% of the genome remaining, 5.7% lay in regions with a LOD score for linkage to the renal disease of below 0 (indicating that it was more likely than not that the observed pattern of segregation was by chance rather than due to linkage at this locus). 3.6% of the genome lay at regions linked with LOD between 0 and 1 and 0.7% were linked with LOD > 3 with the renal disease. This 0.7% comprised one locus on each of chromosomes 1, 2 and 16 (***Figure 3.16-2***).

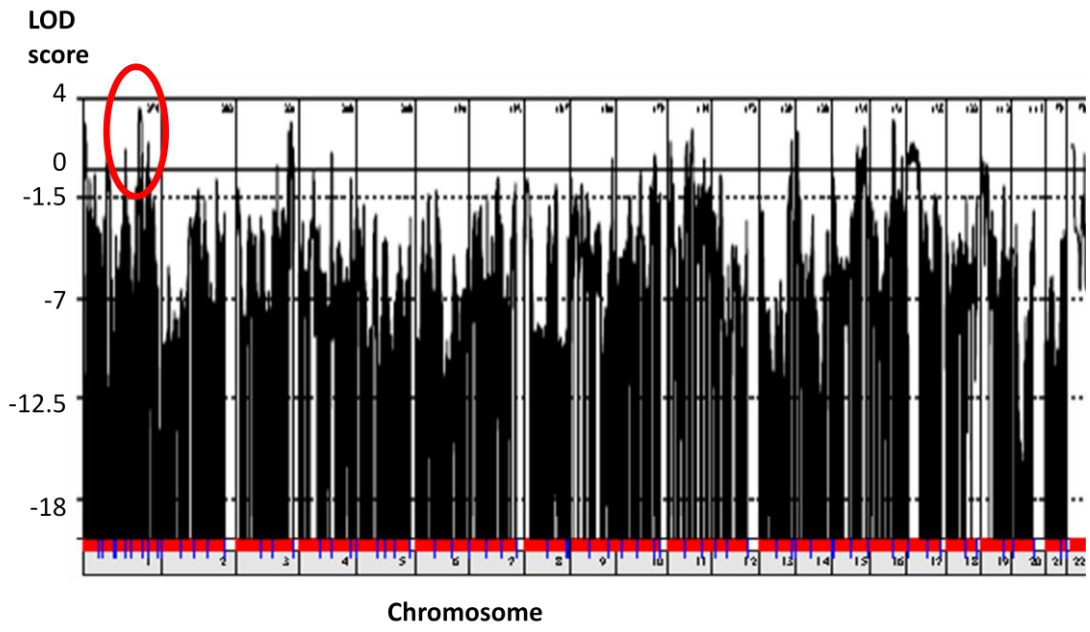


Figure 3.16-2 Genome-wide LOD scores for linkage with renal disease
A single locus (circled) approached the maximum possible LOD score (LOD was 3.4) in families 1 and 2.

Linkage analysis in multiple families generates a LOD score based on the transmission of alleles *within* each family. The presence of linkage with a disease at the same locus in multiple families does not imply that there is identical-by-descent (IBD) inheritance of an allele at this locus. Rather, the LOD score approach was developed specifically *not* to assume this – allowing the analysis of multiple *unrelated* families each with a private mutation at the same locus (i.e. in the same gene) causing a particular disease. Assuming IBD inheritance would prevent simultaneous analysis of multiple such families with a shared phenotype.

However, I considered it likely that there *was* IBD inheritance of the disease-causing mutation for two reasons. Firstly, the phenotype (autosomal dominant inheritance of C3GN) is exceedingly rare and has not previously been recognized, so coincidental identification of 2 families with independent mutations was correspondingly unlikely (note that this argument would not apply if I had started the investigation by screening a very large number of families with kidney disease and *selected* another family with C3GN). Secondly, both families could

trace ancestry to a defined geographical area and the prior probability of a common ancestor was therefore high. It followed that, if linkage at a particular locus in both families was a consequence of the presence of a single disease-causing mutation at that locus, the haplotype surrounding it would be shared, with the size of the shared region dependent (in a statistical sense) on the number of recombinations (i.e. generations) between the families. Therefore, I inspected haplotype plots (inferred by GENEHUNTER and visualised using HAPLOPAINTER) to identify which of these loci was likely to have been inherited from a common ancestor in all affected individuals of both families. This demonstrated that only part of the chromosome 1 linked interval was inherited IBD in both families (***Figure 3.16-3***).

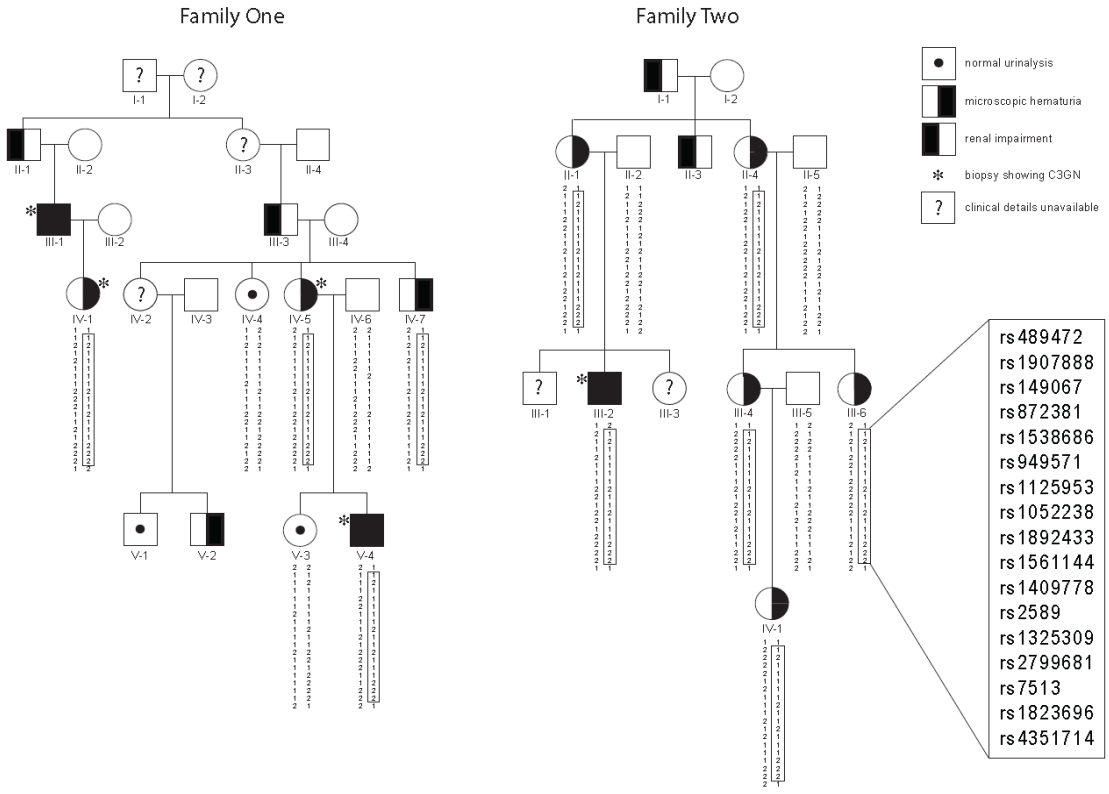


Figure 3.16-3 Identical By Descent (IBD) inheritance
An identical haplotype on chromosome 1 was inherited by all affected members of both family and family 2.

The common haplotype comprised 17 SNPs and encompassed 8.74 cM, a region which included the *CFH/CFHR1-5* gene cluster (***Figure 3.16-4***).

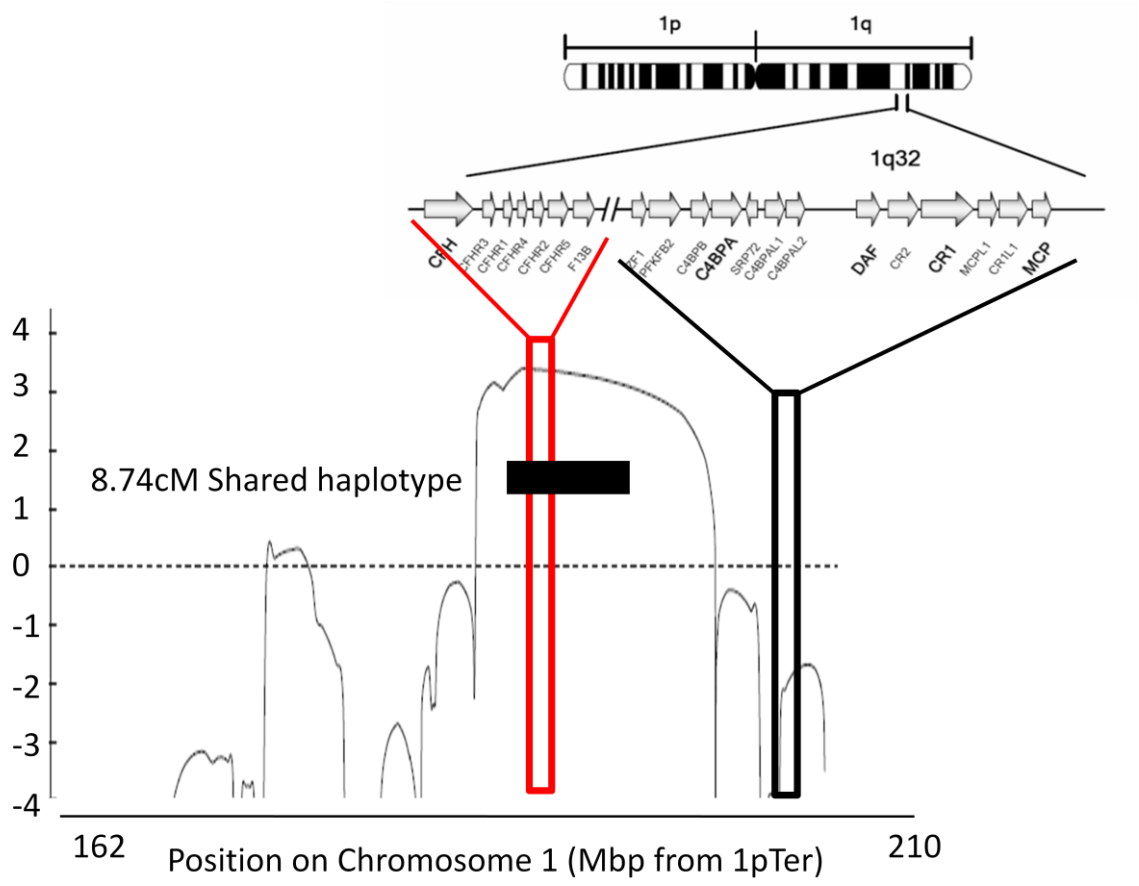


Figure 3.16-4 Linkage, IBD haplotype and genes of the RCA cluster
RCA, Regulators of Complement Activation on Chromosome 1q32

I concluded that the allele (i.e. mutation) responsible for causing the renal disease lay within this region which had been inherited from a common (presumably affected) ancestor by all affected individuals from both families.

3.17 Further investigation of CFH/CFHR1-5 gene cluster

Having amplified and sequenced exons and flanking intronic sequences of *CFH* and *CFHR5,* I considered there to be 2 likely possibilities. The first was that the mutation responsible for causing the disease lay in a different gene in this region (obvious candidates being *CFHR1-4*) and the second was that the mutation was a copy number variation (CNV) in either *CFH* or *CFHR5* which was too large to be identified by single-exon amplification and sequencing.

In order to address this, I first inspected the sequence traces of *CFH* and *CFHR5* from affected individuals and identified heterozygosity for at least one database SNP in each of these genes in at least one affected individual, effectively excluding heterozygosity (or homozygosity) for a whole-gene deletion of either of these genes. Clearly, other types of CNV (for instance whole-gene or sub-whole-gene duplications, hybrid genes and sub-whole-gene deletion events) could not be excluded so simply. I therefore proceeded with direct exon sequencing of other genes in the interval and, in parallel, assays for copy number variation.

3.17.1 *Sequencing of* CFHR1

Since commonly encountered polymorphic deficiency of *CFHR1* is associated with increased risk of disease as a consequence of complement activation, it seemed possible that a mutation in *CFHR1* might cause autosomal dominantly inherited C3GN. Consequently, this gene appeared to be the most attractive of the 'unexplored' genes in the candidate locus. I therefore designed primers in order to amplify and sequence the gene. Because of the high degree of sequence identity present in this region, amplification of the exons of *CFHR1* required PCR primers of high specificity. I designed the primers by inserting the database genomic sequence 800bp flanking CFHR1 into the web-based PRIMER3 program and specifying coordinates of each of the 6 exons. Potential primer sequences generated by PRIMER3 were then aligned with the 360 kilobase-pair genomic sequence encompassing all 6 genes of the *CFH/CFHR1-5* cluster (and 1600 flanking base pairs). Any primers with multiple alignments were rejected. In addition, alignment was attempted at reduced stringency. Any primer which had multiple alignments when 4 or fewer mismatches were allowed were rejected, and primers which had multiple alignments when fewer than 5 mismatches were allowed were rejected unless 2 of the mismatches were at the 3' end (predicting inefficient amplification). Therefore, primers were selected to be selective for the desired amplicon. The exception to this was exon 1, which has 100% sequence identity with exon 1 of *CFHR2.* However, 187 bp downstream of the exon (i.e.

within intron 1) the database sequence for *CFHR1* is C, whereas the corresponding position in *CFHR2* is A (***Figure 3.17-1***).

CFHR1 exon 1

>ref|NG_005811.1|:44969-57890 Homo sapiens complement factor H-related gene cluster (CFHR@) on chromosome 1

CCTCCCAAGTAGCTGGGACTACAGGCACCTACCACCACGCCCAGCTAATTTTT
GTATTTTCAGTAGAGATGGGGTTTCACCATGTTAGCCAGGATGGTCTGAAGT
TACCTCATTCTTAATACCAGTTTTAAGTGGAATTTATAACTAGAACAGTGATG
CTGTTATGTTAAGATGAGGACTGATCACCTTCACTTGCTTGCCTACTGATGTA
GCTGAACTCTTGGCTAGAAAAAAGAAGGGGCTTCCTCTTTCCTCTTCAATGGC
CCATTTCTGAATATTCCAAACTCAGAGACTCAGGGACCACAACAAGGAAATT
GAACAGCTTTTATTTTGCTCAAGTTAATATTACATGATAAACTCAGAGTATTAT
TGTGAAAACACTGATTAGACACTATTTGCTTATTTTGCACAACCCTCCATGAA
CTTTGATGTTTACCACAAAGGACTTTACTAAACTAGCTTCCAGTTAGTACACTG
AAATTCAAAGTCATGCTCATAACTGTTAATGAAAGCAGATTCAAAGCAACA
CCACCACCACTGAAGTATTTTTAGTTATATAAGATTGGAACTACCAAGCATG
TGGCTCCTGGTCAGTGTAATTCTAATCTCACGGATATCCTCTGTTGGGGGAG
AAGGTAAGTTCAAAACAGACCTGAATATTTAGTTCCTTTTTCAGATACATTTAT
CGGTTTTTGTGTGTATGCTTACATATTTTTAAATGAATAAATGGATGAAAATA
TTTTAAATGAGTTATAATATTAATCTATTTTATGGAAATACTTTCTAACATGCA
ATTAGCAGGAAAATAGAATAAAATTAGTTCTCTCCATCCTCTAAGTTGCAAAG
GTAAATGGCCACCAAATAGAGAATGTAAAGAGGAATTAAATGAGGGAAAAC
CTGCTATACTAACGGTGGCAAGGTGAAGTATTAGTGTAATTCTGCATACACTC
TGCAATGCCACTAACTAGAACAATCAATCTCAAAAAGATTTGGCCACCTTCAA
AATGCTTTTTGGATTTTAGTTTCCTAGATTTGATCATTTAAATTTTAAAGGTCA
TTAAAATGGAGGAATTCCCATTTCAATTGAATTATCCTTCAACAGAGGGAAAT
ATATTTTACTGAAACAATAGAAAATATCATTTATATATAAACC

CFHR2 exon 1

>ref|NC_000001.10|:196912434-196928319 Homo sapiens chromosome 1, GRCh37 primary reference assembly

CCTCCCAAGTAGCTGGGACTACAGGCACCTACCACCACGCCCAGCTAATTTTT
GTATTTTCAGTAGAGATGGGGTTTCACCAGGTTAGCCAGGATGGTCTGAAGT
TACCTCATTCTTAATACCAGTTTTAAGTGGAATTTATAACTAGAACAGTGATG
CTGTTATGTTAAGATGAGGACTGATCACCTTCACTTGCTTGCCTACTGATGTA
GCTGAACTCTTGGCTAGAAAAAAGAAGGGGCTTCCTCTTTCCTCTTCAATGGC
CCATTTCTGAATATTCCAAACTCAGAGACTCAGGGACCACAACAAGGAAATT
GAACAGCTTTTATTTTGCTCAAGTTAATATTACATGATAAACTCAGAGTATTAT
TGTGAAAACACTGATTAGACACTATTTGCTTATTTTGCACAACCCTCCATGAA
CTTTGATGTTTACCACAAAGGACTTTACTAAACTAGCTTCCAGTTAGTACACTG
AAATTCAAAGTCATGCTCATAACTGTTAATGAAAGCAGATTCAAAGCAACA
CCACCACCACTGAAGTATTTTTAGTTATATAAGATTGGAACTACCAAGCATG
TGGCTCCTGGTCAGTGTAATTCTAATCTCACGGATATCCTCTGTTGGGGGAG
AAGGTAAGTTCAAAACAGACCTGAATATTTAGTTCCTTTTTCAGATACATTTAT
CGGTTTTTGTGTGTATGCTTACATATTTTTAAATGAATAAATGGATGAAAATA
TTTTAAATGAGTTATAATATTAATCTATTTTATGGAAATACTTTCTAACATGCA
ATTAGCAGGAAAATAGAATAAAATTAGTTATCTCCATCCTCTAAGTTGCAAAG
GTAAATGGCCACCAAATAGAGAATGTAAAGAGGAATTAAATGAGGGAAAAC
CTGCTATACTAACGGTGGCAAGGTGAAGTATTAGTGTAATTCTGCATACACTC
TGCAATGCCACTAACTAGAACAATCAATCTCAAAAAGATTTGGCCACCTTCAA
AATGCTTTTTGGATTTTAGTTTCCTAGATTTGATCATTTAAATTTTAAAGGTCA
TTAAAATGGAGGAATTCCCATTTCAATCGAATTATCCTTCAACAGAGGGAAAT
ATATTTTACTGAAACAATAGAAAATATCATTTATATATAAACC

Figure 3.17-1 Genomic sequences of CFHR1 exons 1 and 2

The sequences in these regions are identical except for 2 nucleotides (red) which do not lie within the coding sequence (bold). Position of primers used to amplify this region are shown in blue.

Therefore, amplification of both exons simultaneously would be predicted to produce apparent heterozygosity for a C/A 'pseudoSNP' (i.e. chain termination with ddC is just as likely as termination with ddA following equal amplification of both exons. This results in overlying curves of reduced amplitude on a sequencing electropherogram, exactly as would be produced by the presence of heterozygosity at a particular base-pair in a diploid organism). I therefore used primers predicted to amplify both exons equally efficiently and used the presence of this pseudoSNP to confirm amplification of exon 1 of both genes. In both patients tested apparent heterozygosity at this position was observed (***Figure 3.17-2***), with all other positions demonstrating homozygosity for the database nucleotide (clearly, heterozygosity in only one of the 2 genes would be predicted to result in 2 overlying peaks, one ¼ the amplitude of the other). This demonstrated firstly no mutations within exon 1 of either gene and secondly that there was not homozygous deletion of either gene in affected individuals. The similar size of

the peaks of the pseudoSNP suggested that there was not a duplication or heterozygous deletion involving one of these exons (but could not indicate if a single CNV involving *both* genes had occurred). Sequencing of all the other exons demonstrated no mutations in the 2 affected members of family 1 sequenced.

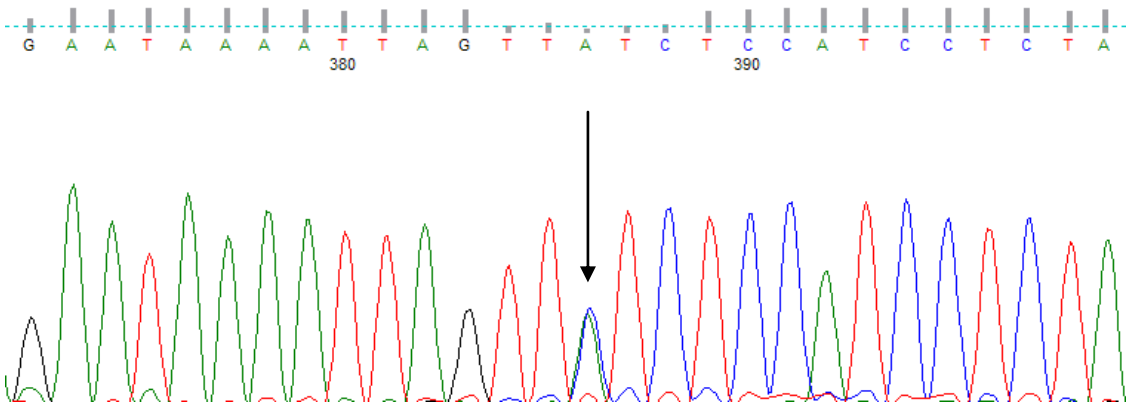


Figure 3.17-2 CFHR1 exon 1 pseudoSNP

In parallel with PCR-amplification and sequencing, analysis of CNV across the locus using 2 independent approaches was performed: Comparative Genomic Hybridisation (CGH) and Multiplex Ligation-dependent Probe Amplification (MLPA) analysis.

3.17.2 Comparative Genomic Hybridisation (CGH)

CGH is a way of detecting changes in copy number by competitive annealing of fluorescently labelled patient and control DNA to a tiled array of probes which are fixed to a glass slide. The ratio of patient to control DNA binding at each position on the slide is related to the strength of fluorescence, detected by a charge couple device. This is then plotted against genomic location to determine the copy number at each location.

Since the genomic region of interest had been defined by the linkage study, I commissioned a targeted array which used 385,000 probes spanning the region inherited IBD in both families. This analysis was performed by Nimblegen in Iceland and the data are summarized in ***Figure***

3.17-3. This analysis revealed heterozygous deletion of *CFHR1* and *CFHR3* in III-2 from family 2, but not in IV-5 from family 1. This is a common polymorphism, present in 14.2% of healthy controls (Dragon-Durey, Blanc et al. 2009) and could not account for the renal disease. Furthermore, since it was absent in an affected individual from family 1 I inferred that individual III-2 in family 2 had inherited it from his unaffected father (i.e. was on the other, unlinked, chromosome). I concluded from this analysis that a significant region of CNV was unlikely to cosegregate with (and hence cause) the disease.

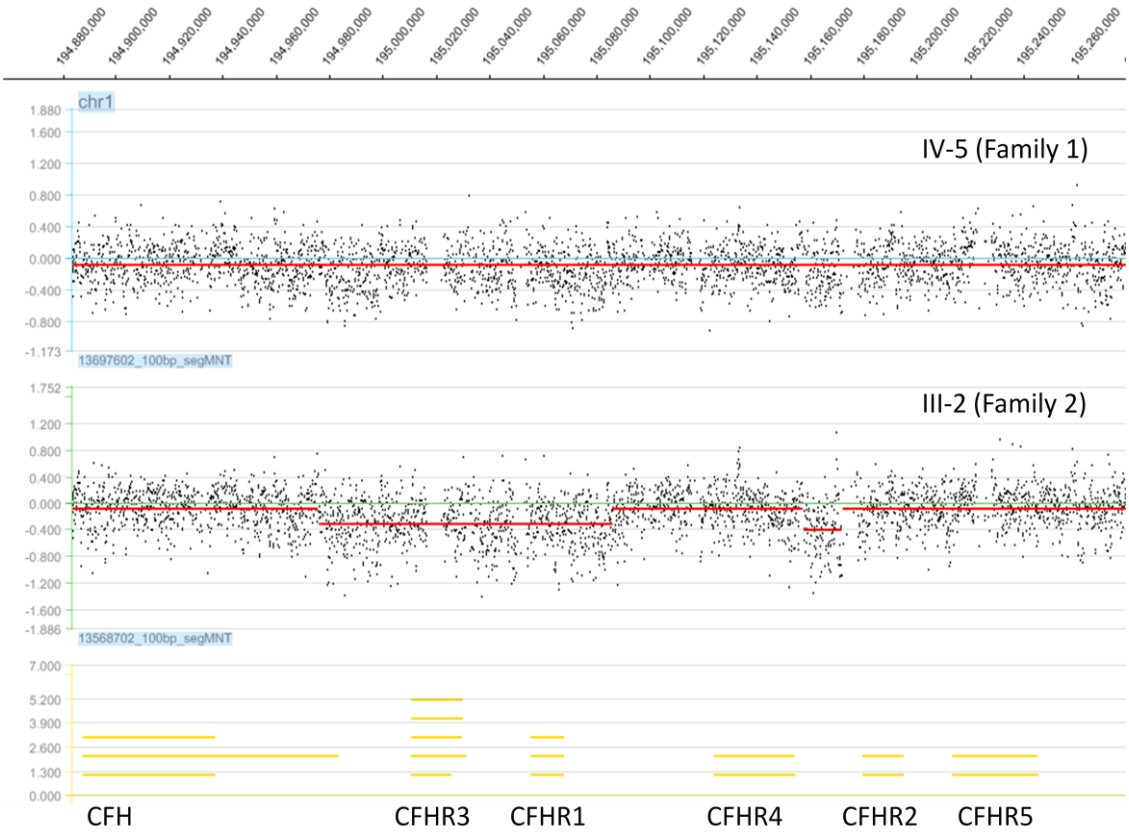


Figure 3.17-3 Comparative Genomic Hybridisation (CGH) array
CGH across the CFH/CFHR gene cluster showing evidence for a deletion of CFHR1 and 3 in III-2 from family 2 only. No other significant copy number variation was identified.

3.17.3 Multiplex Ligation-dependent Probe Amplification (MLPA) analysis

MLPA is a specific, PCR-based method for determining copy number. A pair of probes, complementary to adjacent oligonucleotides, are allowed to anneal to the target DNA. Only when both are annealed (with no mismatch at the 3' end of the upstream probe or the 5' end of the downstream probe) will they be ligated by DNA ligase. Attached to the 5' end of the upstream probe and the 3' end of the downstream probe (i.e. flanking the ligated double probe) are sequences complementary to PCR primers (the forward one of which is fluorescently labelled) which are added to the reaction mixture. Geometric amplification by PCR therefore only occurs of probe-pairs which are exactly complementary (at the ligation point) with target DNA. Since the length of each probe (and its non-complementary tail) can be set experimentally, multiple target sequences can be assayed in a single reaction, utilizing the same amplification primers and separating the amplicons electrophoretically (***Figure 3.17-4***). Commercial MLPA kits are available for regions of the genome in which CNV has previously been considered important, including the *CFH/CFHR1-5* gene cluster and include validated probe pairs with ligation points at sites of unique base-pairs in homologous regions.

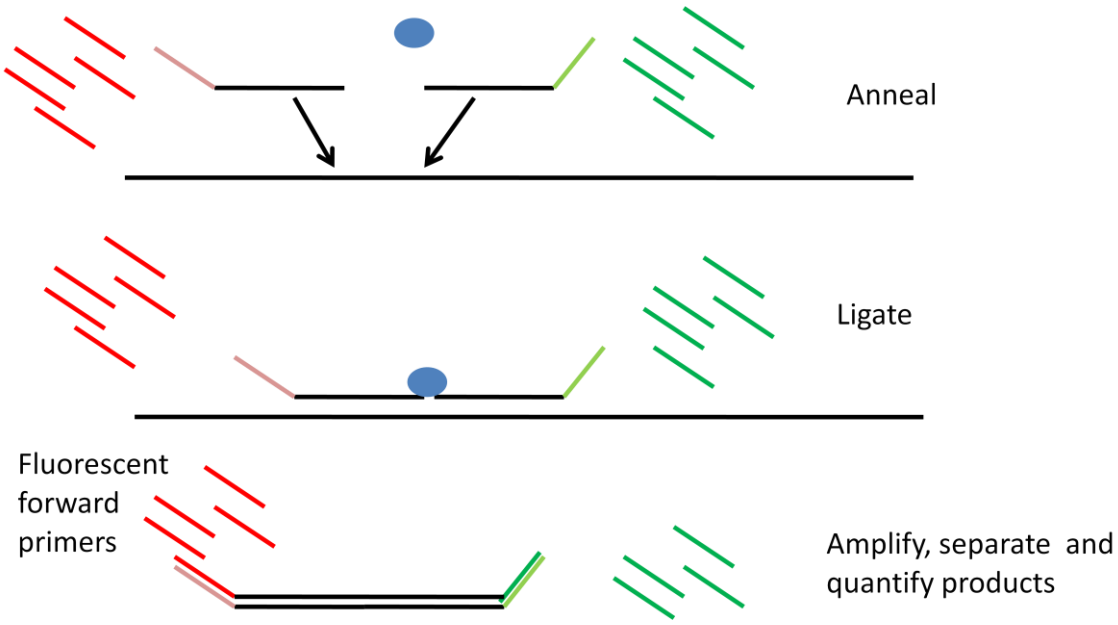


Figure 3.17-4 Multiplex Ligation-dependent Probe Amplification (MLPA) analysis

The gene dosage quotient (DQ) is determined by measuring the signal from the amplicon in question (A) and dividing by the signal from a reference amplicon (B). The ratio of this quotient to the quotient of the same amplicon (a) and same reference amplicon (b) in a control subject is determined to yield a relative copy number (Yau, Bobrow et al. 1996).

Hence: $$DQ = \left(\frac{A}{B}\right) / \left(\frac{a}{b}\right)$$

The experiment was performed by Dr Rubén Martinez-Barricarte in Madrid. Gene dose across the region is indicated in ***Figure 3.17-5***. This confirmed that the heterozygous deletion of *CFHR1*and *CFHR3* which had been identified in the CGH array was present in family 2, and that it had been inherited by III-2 from his father, with the other allele (conferring the disease) inherited from his mother (who also – coincidentally – carried the same common *CFHR1/3* deletion on her other chromosome). This duplication was not present in the individuals tested from family 1.

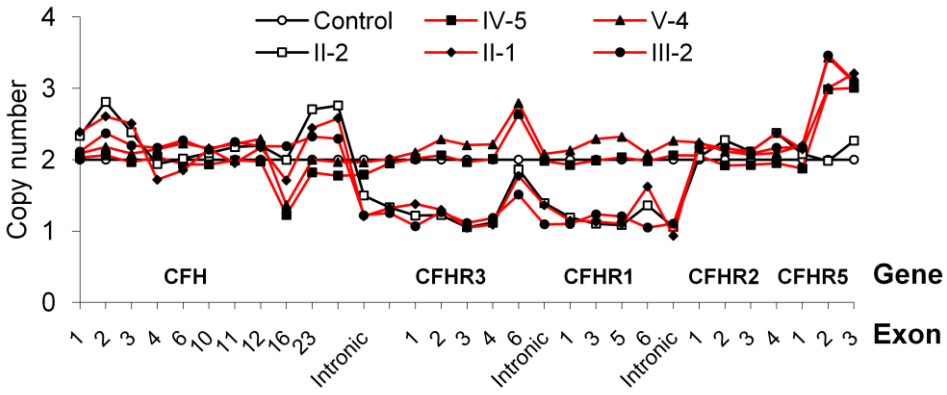


Figure 3.17-5 MLPA analysis in individuals from families 1 and 2
Confirming CFHR1 and 3 deletion in family 2 (in both affected and unaffected individuals) and indicating increased copy number of CFHR5 exons 2 and 3 in all affected individuals tested.

In addition, there was evidence of an increased number of copies of exons 2 and 3 (but not exon 1) of *CFHR5* and this was present in affected individuals from both families, consistent with it residing on the disease-linked haplotype. This variant was not previously described and it seemed possible that this was the mutation responsible for causing the disease.

3.18 Confirmation and delineation of CFHR5 internal duplication

While the MLPA was consistent with a duplication of these 2 exons, there was no evidence of this on the CGH array, even on inspection of the signal from individual probes at this genomic location (***Figure 3.18-1***). In view of this disparity, I attempted to verify (and delineate the extent of) the putative duplication using three different approaches. Firstly by amplification and sequencing of cDNA of CFHR5 obtained from affected individuals, secondly by Southern Blotting of genomic DNA in order to define the exact genomic extent of the duplication and thirdly by Western blotting of patient sera to detect an aberrant protein band.

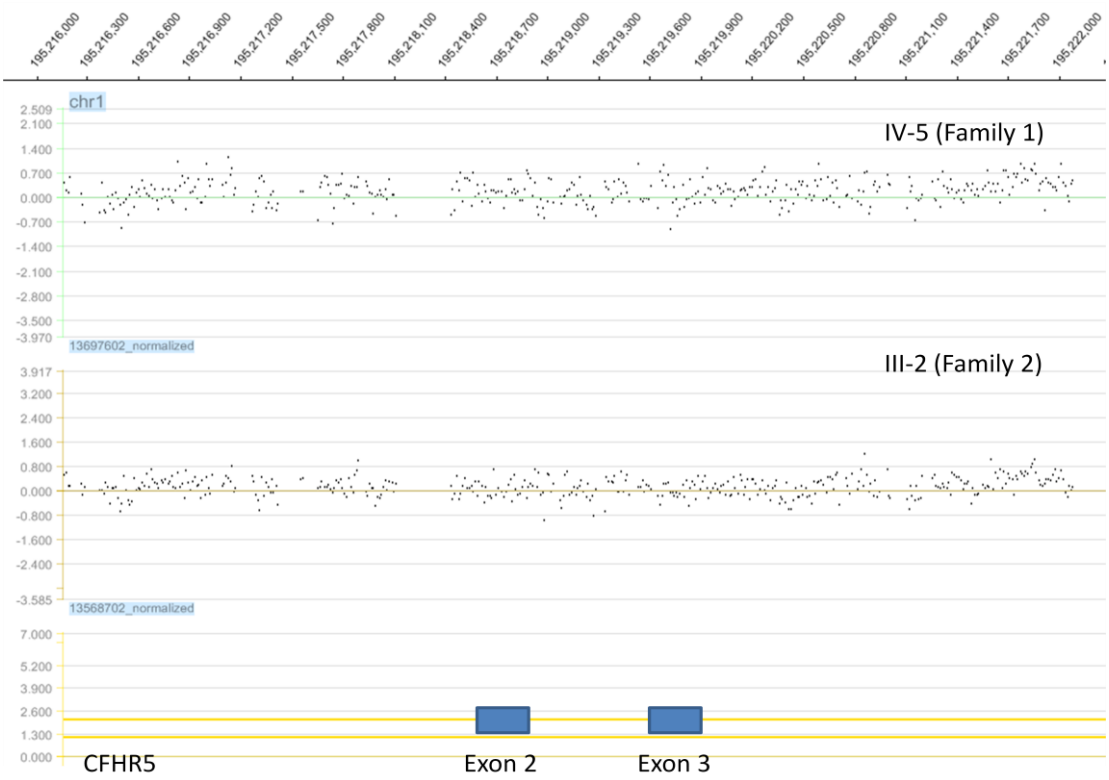


Figure 3.18-1 CGH in detail

Careful inspection of individual CGH probe signal data from affected individuals of both families revealed no evidence of increased copy number at CFHR5 exons 2 and 3.

3.18.1 CFHR5 cDNA sequencing

Consultation of a gene expression database (Su, Wiltshire et al. 2004) predicted low levels of the transcript from cells of peripheral blood (***Figure 3.18-2***). While the transcript is abundantly expressed in normal human liver, I considered that invasive sampling of this organ (i.e. performing a liver biopsy) for the purpose of this investigation would be too high-risk and unpleasant to be ethically justifiable. Nonetheless, I collected fresh blood from individuals IV-5 and V-4 from family 1 and immediately extracted RNA from peripheral lymphocytes. The RNA was retrotranscribed and amplification was attempted using a number of primer pairs for regions of *CFHR5*, but in no case was a product detectable, in either patient or control samples. I concluded from this that the quantity of *CFHR5* transcript in circulating blood was too low to allow this strategy, with the techniques employed, to work.

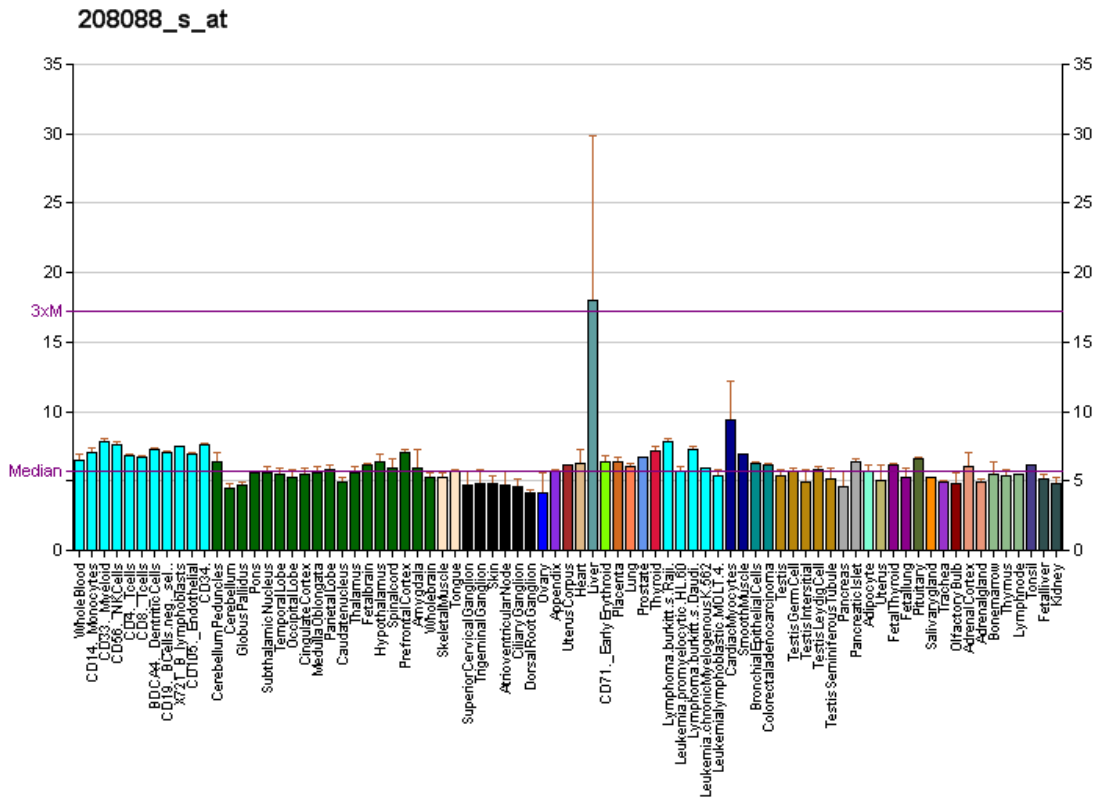


Figure 3.18-2 CFHR5 expression

CFHR5 transcript abundance in publicly available genome-wide expression database, by organ. Data were downloaded from the website specified in (Su, Wiltshire et al. 2004).

3.18.2 CFHR5 Southern Blotting

Consultation of the database sequence surrounding exons 2 and 3 of *CFHR5* indicated that there was an EcoR1 site 439 basepairs (bp) 3' to exon 1 (i.e. in intron 1) and a second site 7934 bp downstream, lying 1958 bp 3' to exon 3 (i.e. in intron 3) (***Figure 3.18-3***). I therefore considered it likely that a duplication involving exons 2 and 3, but not exon 1, would either be flanked by or contain one (but not both) of these sites and that therefore probing a membrane blotted with EcoR1-digested genomic DNA from an affected individual with a radiolabelled PCR amplicon containing exon 2 of *CFHR5* would yield a second band (reflecting the existence of a genomic fragment which contains *CFHR5* exon 2 which migrates at a different rate to the wild-type allele due to altered size). Clearly, there was the possibility that the duplication was large enough to include *both* sites, in which case no aberrant band would be detected.

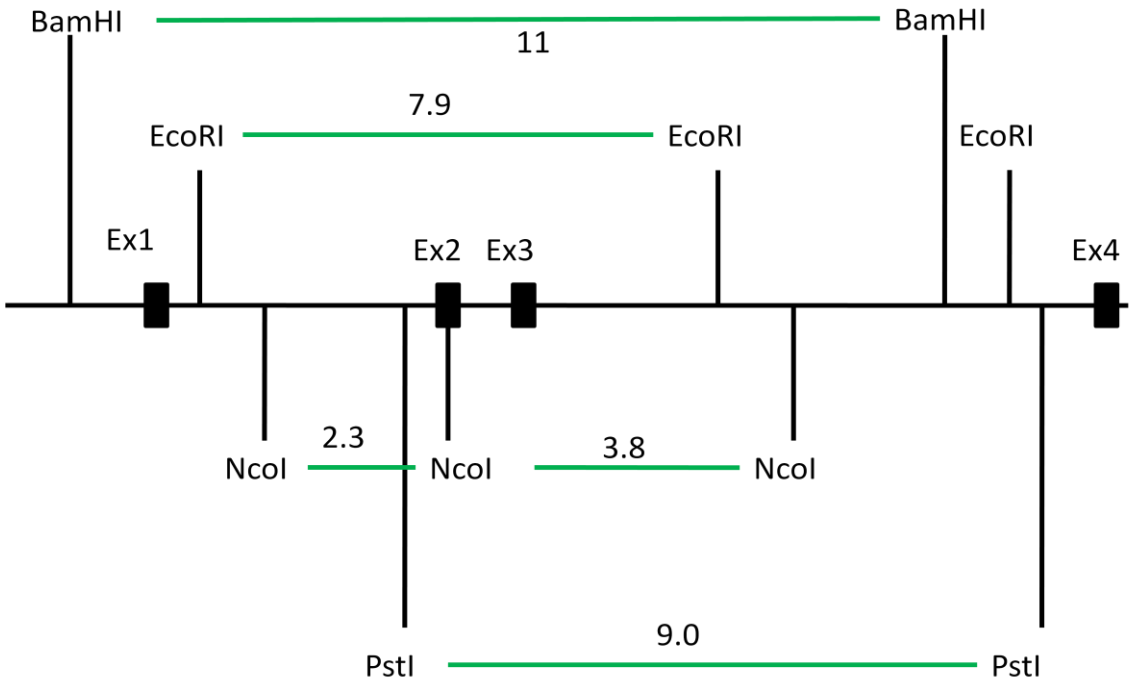


Figure 3.18-3 Restriction sites in relation to exons of CFHR5
Predicted restriction fragment lengths are shown (green lines).

Initial Southern blotting of EcoR1-cut genomic DNA from IV-5 and V-4 demonstrated, in addition to the predicted 7.9 kbp band, a band of 6.3 kbp which was not present in 2 unrelated control individuals (***Figure 3.18-4***). I considered this the crucial experiment proving the existence of genuine CNV cosegregating with the disease (since this allele would have to lie within the shared haplotype in order to be transmitted from IV-5 to V-4). Nonetheless, it was still not clear what the topography of the duplication was.

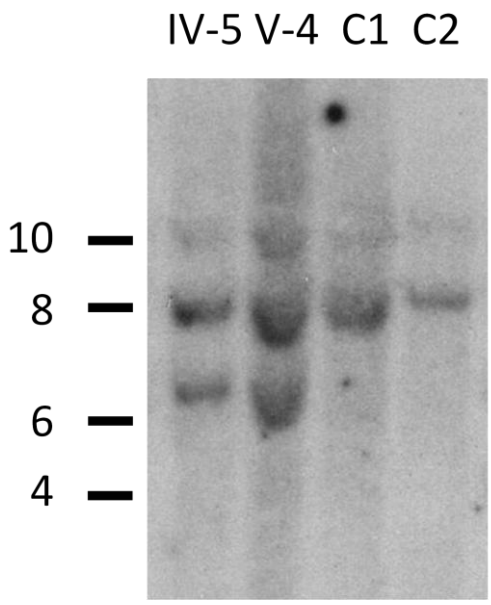


Figure 3.18-4 Southern blot for CFHR5
Southern hybridization of a sequence containing exon 2 of *CHFR5* with genomic DNA cut with *EcoR1* from 2 affected (IV-5 and V-4) and 2 control individuals (C1 and C2) showing an additional 6.3 kbp band in affecteds only.

The MLPA had indicated that the CNV was a duplication (rather than a deletion) and this suggested that one of the EcoR1 sites was included inside a 6.3kbp duplication (if the EcoR1 sites flanked the duplication a *slower* migrating band of the size of the duplication plus 7.9kbp would be predicted). In order firstly to confirm this result in both families and secondly to delineate the topology of the duplication, I repeated the Southern analysis using DNA from multiple affected individuals from both families, digesting each individual with EcoR1 and a second enzyme. These were selected from the restriction map (***Figure 3.18-3***) in the hope of providing the most accurate estimation of the duplication size and position. I selected BamH1 because there was one site 1.2 kbp upstream of exon 1 and a second site 5.2 kbp downstream of exon 2 (and 4.1 kbp downstream of exon 3) – positions I judged likely (although not certain) to flank a 6.3 kbp duplication which included both exon 2 and exon 3 (but not exon 1) of *CFHR5*.

I also selected Nco1 because there was a restriction site within exon 2 and further sites 2.3kbp upstream (in intron 1) and 3.8 kbp downstream (in intron 3). Therefore, a 6.3 kbp

duplication including exon 2 would be likely to result in duplication of 2 of these restriction sites, producing a new band, the size of which may be helpful in determining the insertion point. Finally I included digestion with Pst1 which has one site 278 bp upstream of exon 2 and second site 8.6 kbp downstream. This was likely to result in duplication of either no sites (resulting a product *larger* than the wild-type fragment) or the upstream site only (resulting in an additional band *smaller* than the wild-type band). This would confirm the size of the duplicated region and further limit the possible insertion point. This experiment confirmed firstly inheritance of a 6.3 kbp duplication in affected individuals in both families, secondly that the duplication included one (but not both) Pst1 sites and thirdly that the duplication was flanked by the BamH1 sites (***Figure 3.18-5***). The absence of additional bands resulting from Nco1 digestion implied either that the duplication included all 3 Nco1 sites or else that the distance between the 5'-most duplicated site and the site immediately upstream of the duplication insertion point was either 3.8 or 2.3 kbp.

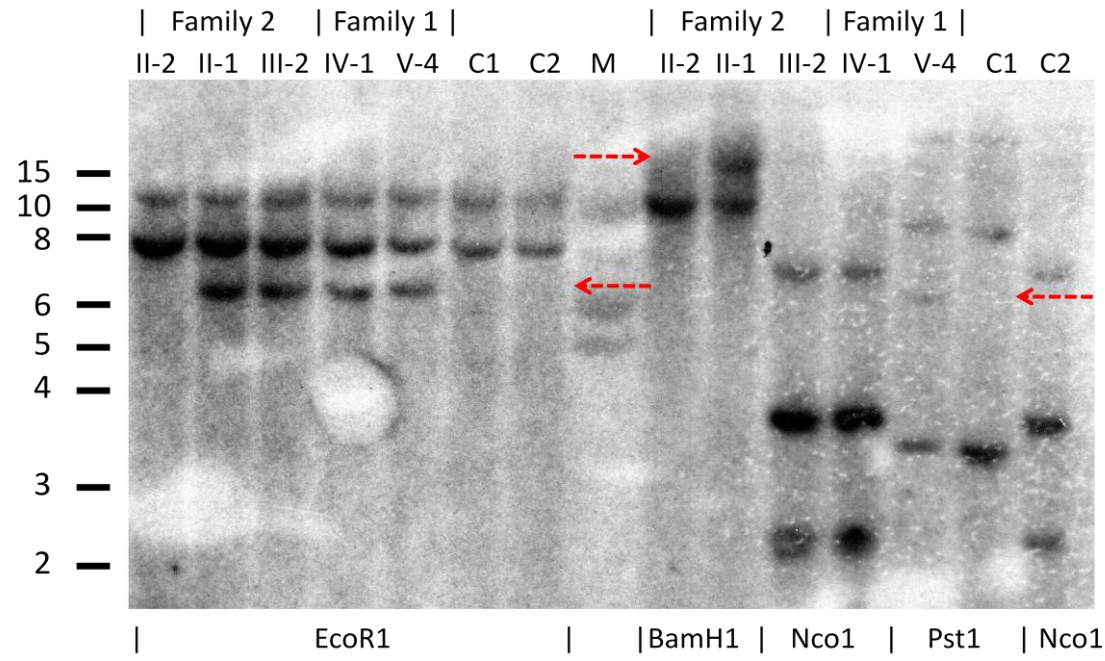
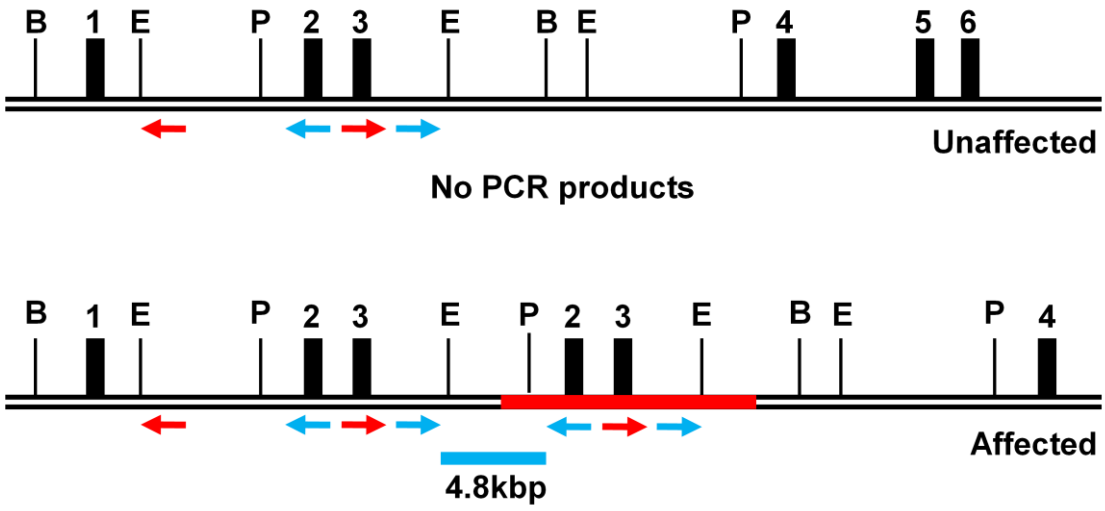


Figure 3.18-5 CFHR5 Southern blot in both families

Southern hybridization of a sequence containing exon 2 of *CFHR5* with genomic DNA from multiple family members and controls digested with various restriction enzymes. Aberrant bands present only in affected individuals are marked (red arrows).

There were, however, 2 possible topologies (with respect to these restriction sites) for the new allele. Either there was duplication of a 6.3 kbp fragment which included the 5' EcoR1 and Pst1 sites, or else the duplication included the 5' Pst1 and 3' EcoR1 sites. In order to determine which of these possibilities was correct I designed primers in which the forward primer was downstream (in the wild type chromosome) of the reverse primer. These would fail to amplify the wild-type sequence but would produce a product if the duplication resulted in a second copy of the reverse primer *downstream* of the forward primer (***Figure 3.18-6***). A product was observed following amplification with primers designed to amplify a duplication involving the 5' Pst1 and 3' EcoR1 sites. This product was 4.8 kbp in size and I performed multiple sequencing reactions (with primers every 700 bp within this product) in order to reveal the sequence of this fragment .



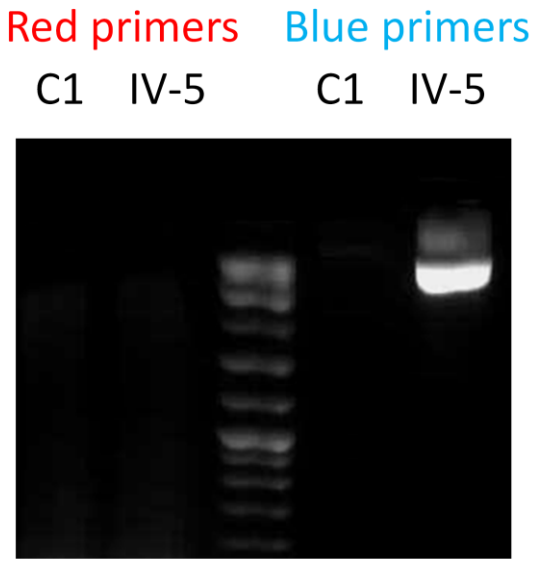


Figure 3.18-6 Selective amplification of duplication
This confirmed its topology with respect to restriction sites and exons. Positions of primers are shown by arrows

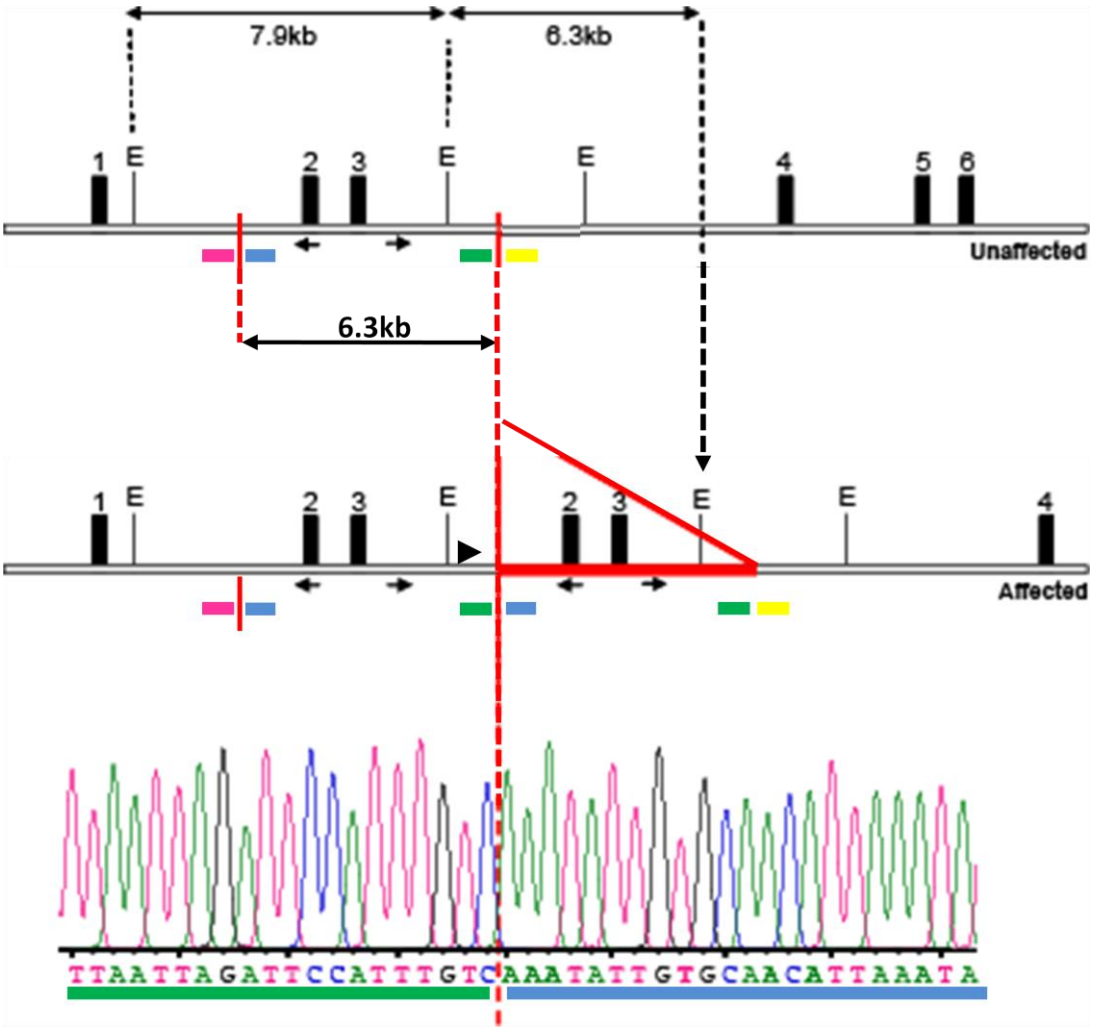


Figure 3.18-7 Exact sequence and position of duplication

Determined by amplification and sequencing of insertion point. Sequence positions are colour-coded and have the order pink-blue-green-yellow and pink-blue-green-blue-green-yellow in the wild-type and mutant alleles respectively.

Comparison of this sequence with the database revealed the exact genomic location of the new sequence and this information allowed me to infer the precise size and location of the entire duplicated region (***Figure 3.18-7***). Interestingly, there was not significant sequence identity between the regions flanking the beginning and end of the duplication, with only a single shared TC pair immediately preceding the start and end of the duplicated region (***Figure 3.18-8***). Blasting the 600bp flanking the start and end of the duplicated section revealed no significant sequence identity, implying that the non-allelelic recombination event which gave rise to the duplication was also non-homologous.

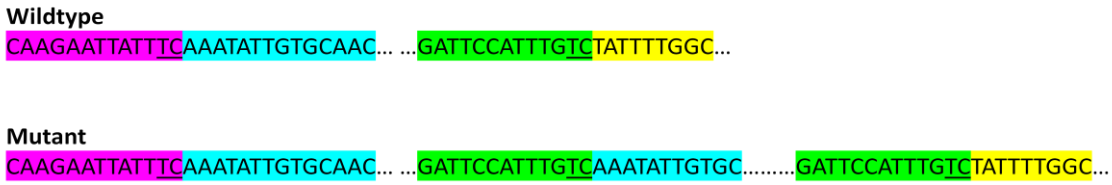


Figure 3.18-8 Local sequences at the exact insertion point of the duplication
Only 2 base pairs (underlined) are duplicated in the sequence immediately prior to the beginning and end of the duplicated section.

3.19 Rapid assay for CFHR5$^{12123\text{-}9}$ allele

Knowledge of the exact genomic sequence of the duplication was important because it allowed me to design a specific PCR reaction which would amplify both the mutant and the wild-type allele at the same time. This reaction was a 1-step diagnostic test for the duplication which would allow its presence to be confirmed in the other members of the family and also allow me to test for the allele frequency in the general population. While it would have been possible to use the presence of a 4.8 kbp fragment as the 'diagnostic' test for the duplication this was suboptimal for two reasons. Firstly, 4.8 kbp is a rather large fragment of DNA to amplify and it is possible that genomic DNA which has become degraded or fragmented would fail to amplify even if the duplication were present; and secondly there was no control, in this reaction, for amplification failure so it would not be possible to determine whether failure to observe a product was due to absence of the duplication in the individual being tested or simply to failure of amplification. I therefore designed a reaction which used 3 primers to produce 2 differently sized products – one as a result of amplification of a wild type sequence in this region, and a second (*smaller*) product as a result of amplification of the mutant sequence. This ensured the test was as sensitive as possible, since small amplicons are more likely to be amplified than larger ones during a PCR reaction, and therefore absence of the small product, in the presence of the larger product, would exclude the presence of the duplication. The positions of these primers are shown in ***Figure 3.19-1***.

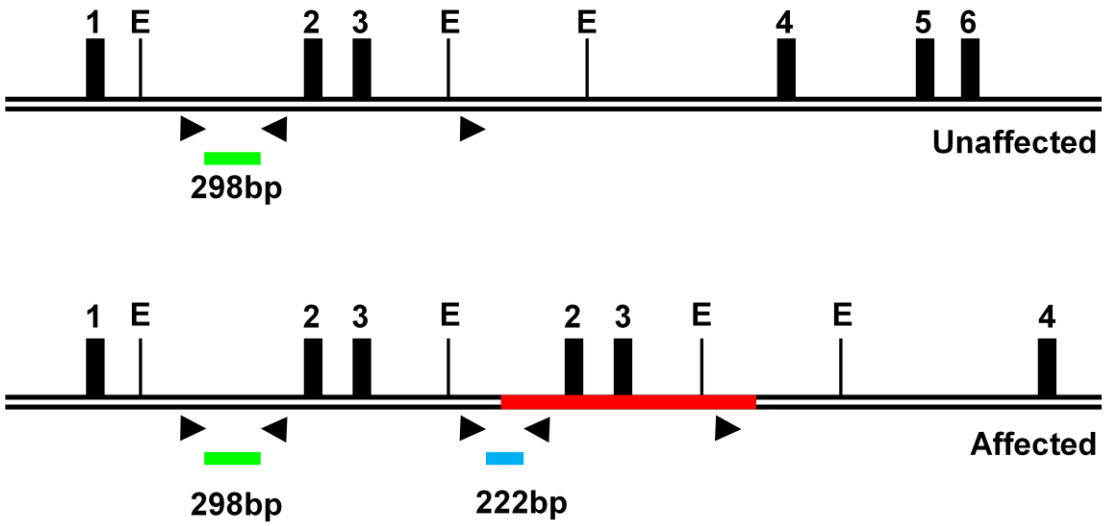


Figure 3.19-1 Duplication-specific 3-primer diagnostic PCR
3 primers (arrowheads) are predicted to amplify a 298-bp fragment in the presence of the wild type allele (green) but an additional 222-bp fragment in the presence of the duplication.

This assay allowed screening of the families and confirmed that the duplication cosegregated with the disease-linked haplotype in both families (***Figure 3.19-2***).

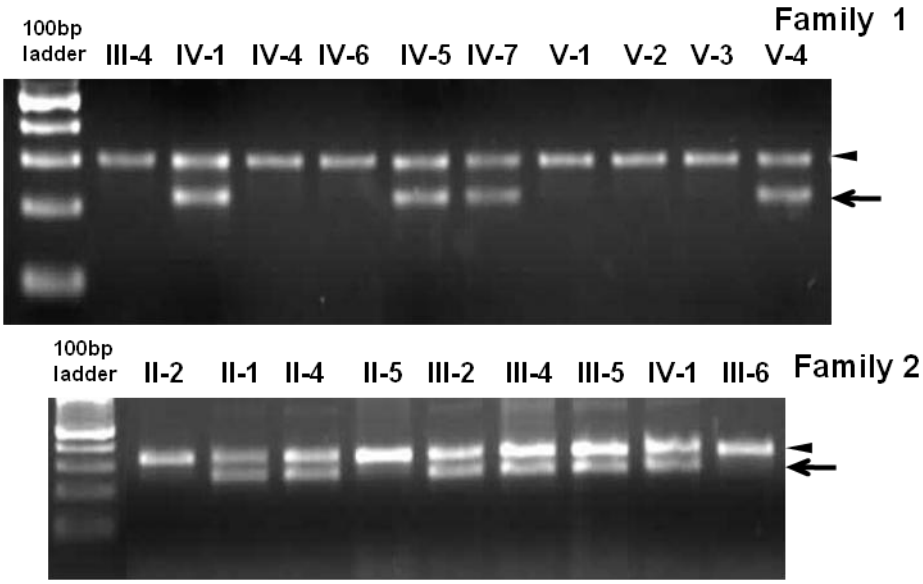


Figure 3.19-2 Family screen by diagnostic 3-primer PCR
Demonstrating the smaller fragment (arrow) produced by amplification across the duplication insertion point in all and only affected individuals from both families. The band produced by amplification of the wild-type sequence (arrowhead) is present in all individuals, indicating that the PCR reaction did not fail.

In addition, I tested 100 individuals from the UK 1958 birth cohort for the presence of the duplication. It was not detected in any of these individuals (***Figure 3.19-3***), implying that this duplication is not a common polymorphism in the general Caucasian (and specifically the UK) population.

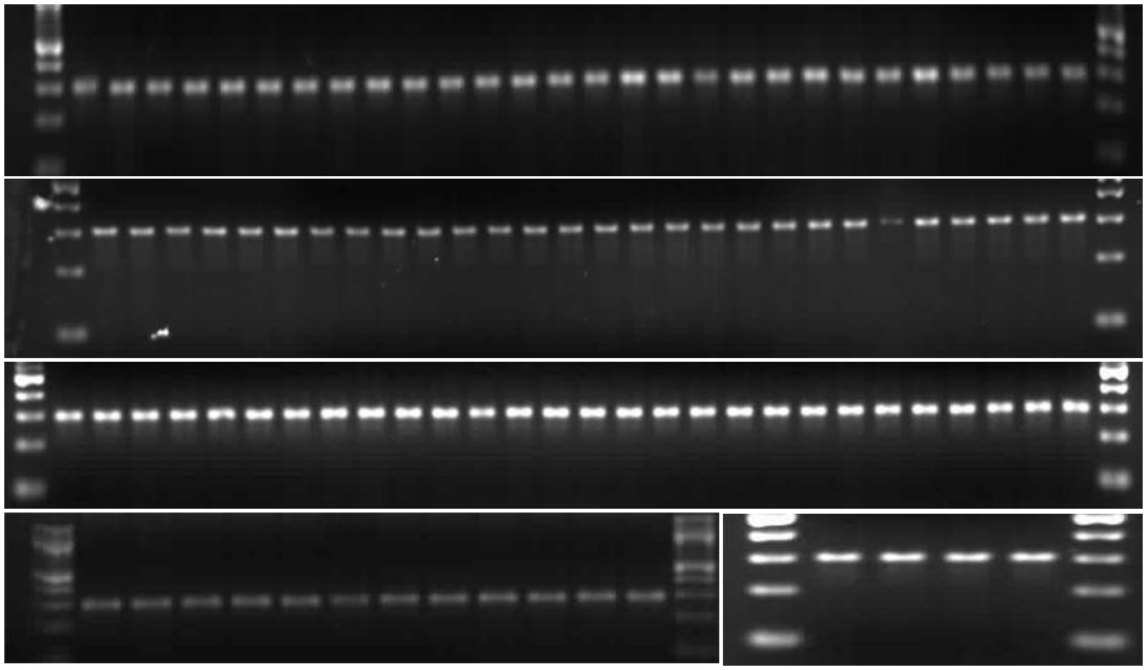


Figure 3.19-3 CFHR5 duplication is absent in UK controls
Only the band produced by amplification of the wild-type allele was detected among 100 individuals from the UK 1958 birth cohort.

3.20 Detection of CFHR5 mutant protein

Inspection of the database sequence indicated that exon 2 is 195 bp (i.e. exactly 65 codons) and exon 3 is 177 bp (i.e. exactly 59 codons) in length. Therefore, duplication of these exons is not predicted to result in a shift of the reading frame of subsequent exons which were therefore predicted to be translated as normal. This would result in a protein of 693 amino acids, 124 amino acids longer than the wild-type CFHR5 protein. The molecular mass of such a protein can be estimated to be ~66 kDa, compared with 54 kDa for the wild type allele (McRae, Duthy et al. 2005). However, there was still the possibility that the mutant protein would be

unable to be exported from the cell or alternatively that it would be too unstable to be detectable in the circulation.

CFHR5 consists of 9 short consensus repeat (SCR) domains, each coded for by 1 of exons 2-9 of the gene respectively. SCRs 1 and 2 (coded for by exons 2 and 3 respectively) are homologous with corresponding SCRs of CFHR1 and CFHR2. SCRs 3-7 and 8-9 share homology with SCRs 10-14 and 19-20 of CFH (see ***Figure 3.10-2***). Therefore, duplication of exons 2 and 3 would be predicted to produce a protein consisting of SCRs 12123456789 (denoted $CFHR5^{12123\text{-}9}$, where the wild-type protein is $CFHR5^{123\text{-}9}$).

Western blotting was performed on patient and control sera by Dr Elena Goicoechea de Jorge using polyclonal antiserum directed against CFHR5. This demonstrated the presence of an aberrant band sharing immunoreactivity with wild type CFHR5 only in individuals with genetic evidence of the CFHR5 duplication. This new band migrated at a rate consistent with a molecular mass of 65 kDa, compared with 54 kDa for the wild-type protein (***Figure 3.20-1***). This was interpreted as most likely resulting from the presence of the mutant protein in detectable quantities in the circulation.

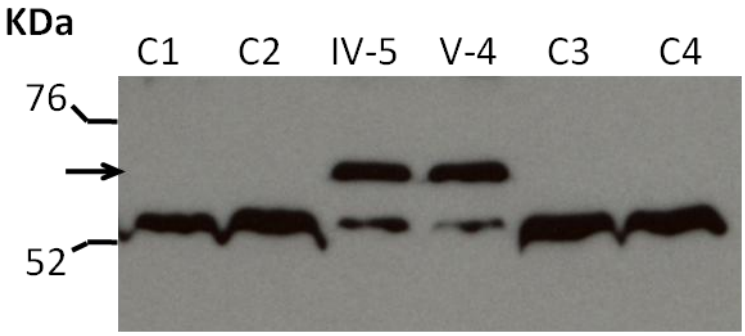


Figure 3.20-1 CFHR5 Western blot of serum
Western blot for CFHR5 of serum from affected individuals (IV-5 and V-4) revealed an additional band migrating more slowly (arrow). This was not seen in controls (C1-4) and is consistent with expression of the mutant allele.

3.21 Discussion

3.22 Copy Number Variation and Human Disease

Copy number variation (CNV) refers to an increased or decreased number of copies of a particular sequence of DNA, which can be of anywhere between a few base pairs or many millions of base pairs in length. It is now known that CNV accounts for a significant amount of the genetic variation between human beings, with some estimates suggesting that approximately twice the number of base pairs of difference between any two individuals being accounted for by CNV compared with SNPs (Korbel, Urban et al. 2007).

3.22.1 How does CNV occur?

CNV usually results from non-allelic recombination, where two previously separated chromosomal segments line up incorrectly and are joined to produce a new sequence. This can occur either in the presence of homology (called non-allelic homologous recombination, NAHR) or in its absence (non-allelic non-homologous recombination). It follows that where a duplication has already occurred this will result in homologous sequences arranged end-to-end on the chromosome (termed a low copy number repeat or LCR). Conventionally, LCRs are defined as regions of >1 kbp sharing >95% sequence identity (Hastings, Lupski et al. 2009) and because of their extensive sequence identity, they provide the substrate for further NAHR events – leading to the development of highly complex genomic regions.

Homologous recombination underlies several mechanisms of DNA repair, where a homologous sequence (presumably most often from the undamaged sister chromosome) is used to repair a damaged sequence (reviewed in (Hastings, Lupski et al. 2009)). Homologous recombination requires sequence identity of 300 bp in mammalian cells (Liskay, Letsou et al. 1987) because smaller (or less perfect) regions of sequence identity are prevented from participating in NAHR by mismatch repair mechanisms which are present in prokaryotic and

eukaryotic cells (Rayssiguier, Thaler et al. 1989; Kim, Krasieva et al. 2002). Novel genetic variants generated by NAHR arise with increased frequency in regions of complex genomic architecture (such as the *CFH/CFHR1-5* gene cluster) (Lupski 1998) and it seems likely that the homology between the different CFH SCR domains and between the *CFH* and *CFHR* genes is a result of non-allelic recombination events over time, and that once a duplication has occurred further non-allelelic recombination events become more likely at this locus due to the proximity of multiple homologous sequences. However, while explaining how CNV can be increased in a region of the genome, NAHR cannot account for the initial generation of CNV and nor can it explain CNVs in which the flanking regions (i.e. the breakpoint) are not homologous.

Non-homologous non-allelic recombination can occur as a consequence of repair mechanisms which do not require homologous sequences, such as non-homologous end joining, where the two ends of a double strand break are ligated – either accurately or with any missing sequence filled in with random nucleotides (Lieber 2008) or by microhomology-mediated end joining, in which a double strand break undergoes end-resection until 5-25 bp homologous sequences are revealed (McVey and Lee 2008). Other non-homologous recombination mechanisms include replication slippage, in which sequence identity exists within an Okazaki fragment[a] (100-200 bp in eukaryotes) and the replication machinery 'slips' between the sequences resulting in small insertions or deletions (Albertini, Hofer et al. 1982); or fork stalling and template switching, in which, during replication, secondary structure in the lagging strand template (or nucleotide shortage) stalls replication leading to a free 3' end of the new DNA molecule which can align with another exposed single-stranded template sequence on another replication fork leading to a duplication, deletion, inversion or translocation (Hastings, Lupski et al. 2009). Interestingly, non-homologous recombination events, although they do not coincide with LCRs or they would be indistinguishable from NAHR events, are also

[a] The discontinuously synthesised piece of DNA on the lagging strand during DNA replication

more common in the areas of complex genomic architecture (such as the CFH/CFHR gene cluster), possibly owing to increased frequency of double strand breaks (Lee, Inoue et al. 2006).

3.22.2 Why does CNV occur?

Although mechanistically CNV appears to occur as by-product of inaccurate DNA repair mechanisms, it has been suggested that in some circumstances CNV may confer adaptive advantages on the host, for example by allowing gene expression levels to be higher than can be achieved purely through transcriptional control (Hastings, Lupski et al. 2009). This view is supported firstly by the observation that genes involved in olfaction, secreted proteins and innate immunity (all with functions relevant to interactions with the environment) appear to be over-represented among regions of CNV and secondly by the observation that human populations which traditionally consume more dietary starch tend to have more copies of the gene for salivary amylase *AMY1* suggesting an adaptive mechanism in response to an environmental challenge (Perry, Dominy et al. 2007). However, it could also be argued that CNV is rapidly purged from the population if it occurs in genes for which precise copy number is important for function and that those genes for which CNV commonly exists are genes for which selection on this basis is less potent (Nguyen, Webber et al. 2008). While it has been suggested that most CNV polymorphisms have been observed at low frequency in the population and that this indicates that they are not tending towards fixation (i.e. are detrimental rather than neutral or beneficial), ascertainment of CNV is not straightforward and a recent analysis has suggested that over 80% of CNV between pairs of humans occurs as a consequence of CNV polymorphisms with allele frequency >5% (McCarroll, Kuruvilla et al. 2008).

CNV can favour increased diversity at a locus in at least 2 ways: firstly by increasing the probability that further CNV will arise and secondly by making it less likely that any mutation which does occur will be detrimental (since a mutation which disables or confers a new

function on one copy of a duplicated gene will have no effect on the sequence of the original gene which will be unchanged). It may be that for genes involved in combating microbial infection, the presence of 'spare copies' of a gene may confer an advantage by allowing more rapid evolutionary change which in turn allows complex organisms to cope with selection by rapidly evolving microbial pathogens. While a possible evolutionary benefit of CNV remains unproven, it has been shown that humans and chimpanzees have more LCRs than mice and rats (Tuzun, Bailey et al. 2004; Dumas, Kim et al. 2007) and it may be that this has allowed faster evolution (over a given number of generations) in primates.

While a potentially beneficial role of CNV is the subject of much speculation, in the *CFH/CFHR1-5* gene cluster CNV is well recognized, common in the general population and is usually associated with *increased* risk of disease. Examples include deletion of *CFHR1* and *CFHR3*; deletion of *CFHR1* and *CFHR4* (Schmid-Kubista, Tosakulwong et al. 2009)*;* and gene conversions involving *CFH* and *CHFR1* (Venables, Strain et al. 2006). As discussed above, there is mounting evidence for pathogenic role for deletion of *CFHR1:* individuals with homozygous (or compound heterozygous) deletions of this gene are overrepresented in collections of patients with aHUS compared with collections of healthy individuals (although the pathophysiology is not fully understood). However, although this is consistent with a detrimental effect of these variants, it must be remembered that since the introduction of antimicrobial therapy and modern medicine the cost-benefit balance of genetic changes which combat microbial infection in the human population have shifted substantially towards the cost side: individuals who lack variants conferring increased resistance to microbial infection can often be protected pharmacologically more effectively than those with variants which are protective against infection but at the same time increase the risk of autoimmune attack.

While gene dosage effects such as these and others (Aitman, Dong et al. 2006; Schaschl, Aitman et al. 2009) are well recognized in determining risk of 'sporadic' (i.e. non-Mendelian)

disorders, the proportion of highly penetrant inherited (i.e. Mendelian) diseases which result from CNV of the sub-gene scale, compared with longer-recognized mechanisms of variation identifiable by PCR-amplification and sequencing (such as small insertions and deletions or missense and nonsense single base-pair substitutions), has yet to be ascertained. Studies in the Mendelian disorders phenylketonuria and long QT syndrome indicate that whole exon duplications or deletions currently account for a small proportion of such cases (Gable, Williams et al. 2003; Eddy, MacCormick et al. 2008). As more powerful new techniques (including whole-genome sequencing) are used to investigate more people it is probable that the full extent of CNV, as a mode of inter-individual variation within a species, will become appreciated.

3.23 "CFHR5 nephropathy"

The results presented here demonstrate the existence of a mutant, expressed CFHR5 variant present in individuals from 2 ostensibly unrelated families from the Troodos region of Cyprus affected with synpharyngitic macroscopic haematuria, C3GN and renal failure. These findings therefore potentially define a new clinical and pathological entity (designated "CFHR5 nephropathy") but further information was needed – firstly to confirm the pathogenic nature of the CFHR5 mutation; secondly to elucidate the mechanism by which it causes disease; and thirdly to define the prevalence and geographical distribution of the disease. This is the subject of Chapters 4 and 5.

Chapter 4: CFHR5 functional studies

4.1 Introduction

In the previous chapter I described the identification of a novel, expressed variant of CFHR5 (an in-frame duplication of exons 2 and 3, coding for a protein denoted CFHR5$^{12123\text{-}9}$) which cosegregates with renal disease in 2 ostensibly unrelated families. Histological examination of kidney biopsy specimens from affected members of both families revealed C3 glomerulonephritis, in which complement C3 is deposited in the kidney in the absence of any deposition of immunoglobulins or C1q. This suggested a defect of complement alternative pathway regulation.

As discussed in the previous chapter, the physiological function(s) of the CFHR proteins are not well understood and, prior to the identification of a Mendelian trait cosegregating with a CFHR5 mutation, the evidence of an important biological role for any of these proteins was somewhat circumstantial. However, recent work has revealed genetic evidence that complete deficiency of CFHR1 is a risk factor for the development of antibodies directed against CFH which in turn predispose to atypical haemolytic uraemic syndrome (aHUS) (Abarrategui-Garrido, Martinez-Barricarte et al. 2009; Moore, Strain et al. 2010). This suggests that at least some CFHR proteins have a non-redundant biological role in humans.

4.1.1 Genetic epidemiology of CFHR5

Three studies in which the question of whether variation in CFHR5 could contribute to risk of disease have been published to date, investigating dense deposit disease (DDD), haemolytic uraemic syndrome (HUS) and age related macular degeneration (AMD). (Abrera-Abeleda, Nishimura et al. 2006) showed in 22 patients affected with DDD and 131 controls (unaffected by either DDD or age related macular degeneration, AMD) that the minor alleles of 2 SNPs (rs9427661 and rs9427662, located in the promoter region of CFHR5) were significantly less common in the patients (1 out of 44 compared with 28 out of 206 alleles reported, $p = 0.033$ for both SNPs). In addition, the minor allele of a non-synonymous coding SNP, rs12097550

(which codes for P46S, a Proline to Serine substitution at codon 46 in exon 2) was significantly more common in people with DDD compared with controls (3 out of 44 compared with only 1 out of 206 controls, p = 0.00023).

Monteferrante et al (Monteferrante, Brioschi et al. 2007) observed that uncommon genetic variants of CFHR5 were more frequently identified in people with HUS compared with controls. In 80 blood donor controls and 45 patients with HUS and no mutations identified by direct exon sequencing of CFH, PCR amplification and sequencing identified a number of polymorphisms of CFHR5. Although 17.5% patients compared with only 5% controls (p = 0.02) harboured CFHR5 variants, all of these variants were present in at least one unaffected individual. The authors conclude that the data were consistent with variation in CFHR5 playing a 'facilitative but not causative' role in the pathogenesis of HUS.

Importantly, neither of the above studies employed techniques capable of detecting copy number variation and so would not have detected the $CFHR5^{12123\text{-}9}$ allele or other similar variants. In addition, the extent to which the associations reported in both the above studies could have arisen as a consequence of linkage disequilibrium (where an undetected allele of the same or a nearby gene is inherited more often than would be expected by chance with the observed allele of CFHR5) rather than owing to a functional effect of the observed variants was not addressed. Abrera-Adeleba et al conclude in their paper that "functional studies would be required to resolve this question."

More recently, an additional study did not identify (by exonic amplification and sequencing) any pathogenic mutations in CFHR5 in a cohort of 639 unrelated patients with AMD, and the authors argue against an important contribution of variation in CFHR5 in this condition (Narendra, Pauer et al. 2009). Clearly, this study does not prove that CFHR5 does not play a role in protecting the retina against damage, but it does show that SNPs in *CFHR5* are not a

common contributor to genetic risk of this disease. The methods reported by these investigators would not be expected to detect CNV within *CFHR5*.

The identification of a highly penetrant Mendelian disease linked to CFHR5 therefore provides a completely new line of evidence that CFHR5 plays an important role in humans. The aim of the work presented in this chapter was to identify whether duplication of SCRs 1 and 2 (predicted to result from duplication of exons 2 and 3 of the gene) results in any detectable change in the known biochemical functions of the CFHR5 protein. It should be emphasised that, in the absence of a complete understanding of the function of the normal protein, only some of those biochemical properties (related to the known functions of CFH) which had been identified in previous work, were interrogated.

The work presented here was the result of a collaborative venture with Dr Elena Goicoechea de Jorge and Dr Matthew Pickering who performed a substantial number of the experiments described.

4.1.2 Established properties of CFHR5

Since its identification in 2001 as a circulating plasma protein, relatively little work has been published describing the function(s) of CFHR5, and has mostly interrogated the biochemical interactions of CFHR5 with reference to CFH – only those properties known to be displayed by CFH have been tested. It has been shown that CFHR5 is almost invariably present when C3 is deposited in glomerular tissue and also that it is able to bind C3b *in vitro* (McRae, Cowan et al. 2001; Murphy, Georgiou et al. 2002). Subsequently, a number of experiments have been performed by McRae et al (McRae, Duthy et al. 2005) indicating some biochemical properties of recombinant CFHR5, including cofactor activity, heparin binding, binding to C-reactive protein and lipoprotein association.

4.1.2.1 Cofactor activity

Recombinant CFHR5 was shown to be able to act as a cofactor for Factor I-mediated cleavage of C3 in the fluid phase, and this activity is dose-dependent. Whether this reflects a significant function of CFHR5 *in vivo* remains uncertain for two principal reasons: Firstly, the concentration of CFHR5 required in this experiment was 8 μg/ml whereas CFH displayed detectable cofactor activity at a concentration of 0.4 μg/ml. Since CFHR5 circulates at levels of 3-6 μg/ml and CFH is present at concentrations of approximately 500-800 μg/ml it would be predicted that, in the healthy human circulation, CFH is in massive excess with respect to its function as a cofactor for Factor I mediated cleavage of C3. Secondly, genetic deficiency of CFH in humans is associated with C3 activation and reduced levels of circulating C3 (Boyer, Noel et al. 2008; Montes, Goicoechea de Jorge et al. 2008), implying that the ability of CFH to regulate C3 in the circulation is *not* redundant in the presence of CFHR5. It therefore remains to be proven whether this property of CFHR5 is physiologically important or if it simply represents a vestige of a common evolutionary origin with CFH.

4.1.2.2 C3 convertase inhibition

Addition of factor D to a mixture containing C3, C3i, Factor B and magnesium chloride results in the generation of C3a, which can be measured by ELISA. Addition of CFHR5 to this mixture produced a dose-dependent reduction in C3a production over the range 8-80 μg/ml, consistent with inhibition of the C3 convertase. CFH exhibited a similar effect, although of greater magnitude. For the same reasons as those discussed above with respect to cofactor activity, it is unclear whether C3 convertase inhibition in the circulation is a physiologically important function of CFHR5. Importantly, CFHR5 was found not to reduce the rate at which factor B is recruited by immobilised C3b in the presence of properdin, factor D, $NiCl_2$ and NaCl, suggesting that CFHR5 does not have decay accelerating activity or C3 convertase inhibitory activity at a surface (see Chapter 3).

4.1.2.3 Heparin binding

CFHR5 was shown to bind to solid phase and column-immobilised heparin. Notably, a higher NaCl concentration was required to elute CFHR5 than CFH, suggesting a *higher* affinity for heparin. The investigators went on to use recombinant CFHR5 fragments comprising either SCRs 1-4 or SCRs 3-7 to show that, while CFHR5^{34567} required NaCl concentration of 250mM for elution from heparin columns, CFHR5^{1234} was detected in the flow-through and wash fractions, indicating that none of domains 1-4 are sufficient to confer detectable affinity for heparin in this experimental setup.

4.1.2.4 Binding to C-reactive protein

Recombinant CFHR5$^{1-9}$ was also shown to bind to immobilised C-reactive protein (CRP), a property inhibited by addition of heparin. Binding was not observed using the CFHR5^{1234} fragment, however significant binding was exhibited by the CFHR5^{34567} fragment (albeit at a reduced level compared with the complete CFHR5$^{1-9}$ protein).

4.1.2.5 Association with lipoprotein

CFHR5 was also detected in high (but not low or very low) density lipoprotein purified from human plasma.

Together, these functional data provide good evidence that CFHR5 can co-localise with C3 *in vivo* and suggest a mechanism by which CFHR5 can target to host surfaces (owing to high affinity for polyanionic moieties such as heparin found there) but leave unanswered the question of what the primary functions of CFHR5 are. In addition, elucidation of these functional properties has allowed researchers to propose the pairing of domains of CHFR5 with their likely homologous domains of CFH (***Figure 4.1-1***).

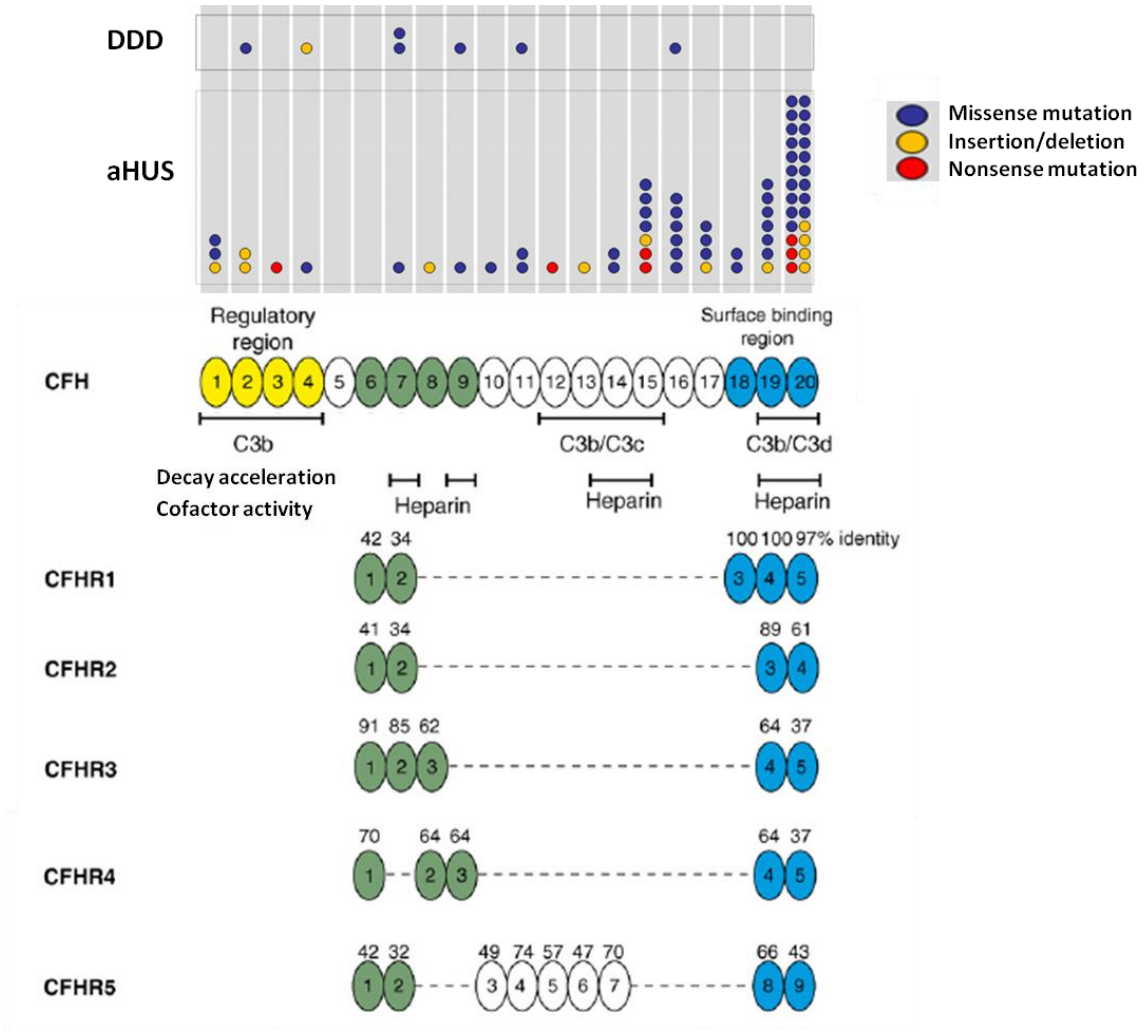


Figure 4.1-1 Proposed homologous domains of CFH and CFH-Related proteins
Positions of CFH mutations associated with dense deposit disease (DDD) and atypical haemolytic uraemic syndrome (aHUS) are shown. Adapted from (Jozsi and Zipfel 2008) and (Boon, van de Kar et al. 2009).

As discussed in Chapter 3, aHUS is believed to result from defective complement regulation at host surfaces and is commonly associated with mutations in the C-terminal 2 domains of CFH. This, combined with the absence of any reported association between mutations in the (functionally) homologous domains of any of the CFHR proteins and aHUS, has two corollaries: firstly, normal levels of functional CFHR proteins are *not* sufficient to protect the host from aHUS in the presence of abnormalities of the C-terminal SCRs of CFH; and secondly that functionally normal levels of CFH *are* sufficient to prevent aHUS. In support of this, and despite extensive screening, the only CFHR alleles which have been shown to be associated with aHUS

are those which are also associated with reduced CFH activity – either owing to the development of autoantibodies against CFH (Moore, Strain et al. 2010) or as a consequence of a likely dominant negative effect on CFH (CFHR1*B in (Abarrategui-Garrido, Martinez-Barricarte et al. 2009).

In summary, while the known biochemical properties of CFHR5 seem to most closely resemble those physicochemical properties which are necessary in CFH to prevent aHUS (i.e. Factor I cofactor activity, C3 convertase inhibition and heparin binding), these properties have not been shown, in any CFHR protein, to be necessary to prevent disease. It can therefore be concluded that, while the biochemistry is partially understood, the physiological role of CFHR5 remains almost completely unknown. Interestingly, none of the published experiments interrogate whether CFHR5 is able to play a role in regulating the terminal complement pathway, although recent work has shown that CFHR1 is able to inhibit C5 convertase (and hence terminal pathway) activity (Heinen, Hartmann et al. 2009).

4.1.3 How might CFHR5$^{12123\text{-}9}$ cause disease?

In principle, an autosomal dominant disease can result from a mutation by at least 4 different mechanisms:

1. **Loss-of-function**: mutations resulting in the production of insufficient quantity of functional protein (i.e. haploinsufficiency) – for example mutations in *CBFA1* which result in cleidocranial dysostosis.
2. **Gain-of-function**: inheritance of an allele for which the protein has either impaired degradation and is constitutively more active (for example impaired breakdown of the epithelial sodium channel due to a mutation in the *ENaC* gene in Liddle syndrome) or else has acquired new physico-chemical properties leading to a new 'function' of the protein which itself causes disease (for example *transthyretin* mutations in familial amyloidosis)
3. **Dominant negative**: the mutant allele is able to impair the function of the wild-type allele resulting in disease – an example of this is Marfan syndrome resulting from inheritance of certain defective *fibrillin* alleles
4. **Somatic second hit**: germline loss of one allele results in complete absence of the functional protein when a mutation (or deletion) occurs in the *other* copy during mitosis in a somatic cell. While autosomal dominant clinically, these disorders can be regarded as recessive on a cellular level – examples include inherited tumour

syndromes, such as mutations in *STK11* in Peutz Jeghers Syndrome or *VHL* in von Hippel Lindau disease

CFHR5 nephropathy appeared most likely to result from either a loss-of-function or a dominant negative mutation in CFHR5, although a gain-of-function mutation could also be responsible.

Understanding the mechanism by which the observed duplication in CFHR5 might cause renal disease is potentially important for two main reasons. The first is in optimising the approach to treating affected individuals: in a simple haploinsufficiency (i.e. loss-of-function mutation) model, it would be predicted that administering wild-type protein (either a pure preparation or as a plasma infusion) to affected individuals may be sufficient to delay the progression of renal impairment. However, if the mutation had a dominant negative or gain-of-function effect then a rational treatment approach would involve removal or neutralisation of the circulating mutant protein – an approach likely to present significantly greater barriers to success than simply supplementing the quantity of the circulating normal protein.

In addition to informing rational design of treatment strategies for individuals CFHR5 nephropathy, it is possible that understanding the mechanism of the disease might provide significant insight into the role of CFHR5 in the normal human kidney, and also its role in other diseases. A disease resulting from haploinsufficiency or dominant negative effects would imply that CFHR5 plays an important physiological role in protecting the kidneys from complement-mediated attack. It follows that supplementing CFHR5 activity in individuals with MCGN as a result of other underlying conditions (such as infection, autoimmunity or IgA disease) may provide a novel therapeutic avenue for the future. A gain-of-function mutation would provoke a more circumspect interpretation of the role of wild-type CFHR5.

4.2 Properties of CFHR5$^{12123\text{-}9}$

4.2.1 Predicted effects of SCR 1 and 2 duplication

Knowledge of the nucleic acid sequence of the mutant allele allowed prediction of the amino acid sequence of the mutant protein. In addition to predicting the molecular mass (by summation of the molecular mass of each amino acid), knowledge of the amino acid sequence allows estimation of the overall charge (and pI) of a protein, using the following equation:

$$Z = \sum_i N_i \frac{10^{pKa_i}}{10^{pH} + 10^{pKa_i}} - \sum_j N_j \frac{10^{pH}}{10^{pH} + 10^{pKa_j}}$$

where N_i is the number and pKa_i the pKa values, of the N-termini and the side chains of the Arginine, Lysine, and Histidine residues; and N_j and pKa_j are the number and pKa values, of the C-termini and side chains of the Aspartate, Glutamate, Cysteine and Tyrosine residues. This equation can be solved for any peptide chain at a given pH using an online calculator (for example at www.innovagen.se) and can also be used to generate the predicted pI or isoelectric point (i.e. the pH at which the charge is zero). While wild-type CFHR5$^{123\text{-}9}$ has a predicted charge of 1.8 at pH 7.0 (and a pI of 7.3), CFHR5$^{12123\text{-}9}$ has a predicted charge of -3.2 at pH 7.0 and a predicted pI of 6.6 (see ***Figure 4.2-1***).

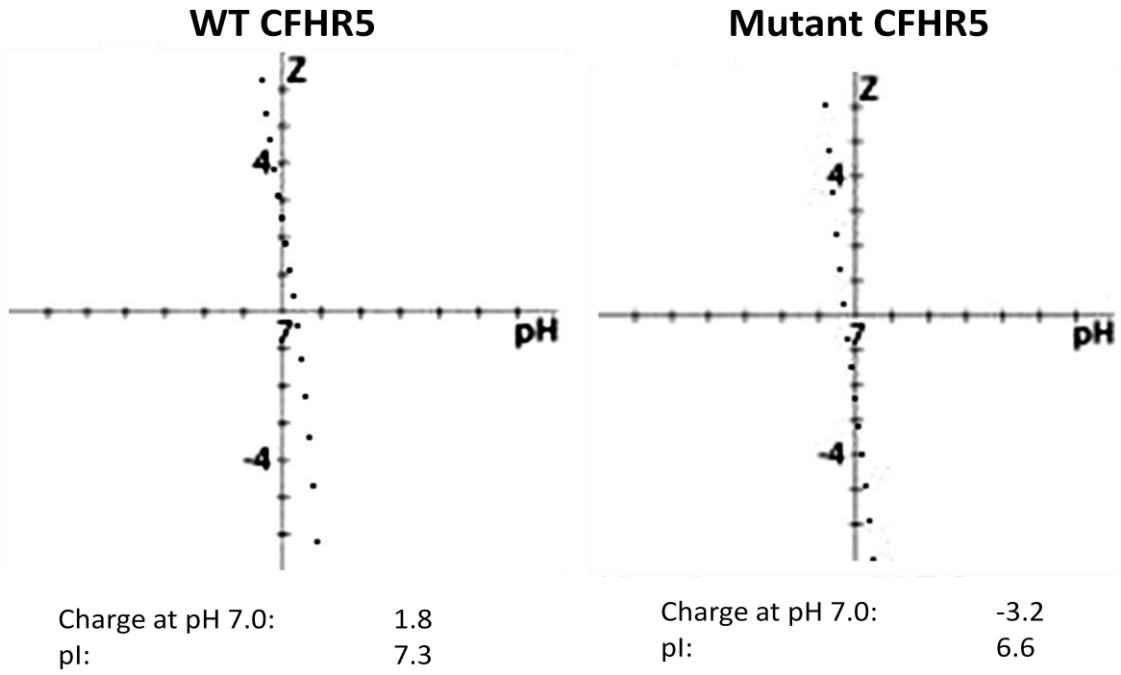


Figure 4.2-1 Predicted charge and pI of wild type (WT) and mutant CFHR5 proteins

This suggests that, while the wild type protein is likely to be positive or uncharged at a physiological pH, duplication of SCRs 1 and 2 is predicted to produce a negatively charged protein – potentially reducing the overall affinity of the protein to negatively charged host surfaces. However, it is important to note that this method of modelling is relatively crude: for instance, it assumes each residue is independent and takes no account of the presence of post-translational modifications or disulphide bridges (of which there are at least 18 in CFHR5[123-9] and 22 in CFHR5[12123-9]).

In addition to the relatively large (from a physico-chemical point of view) change predicted by duplication of 2 whole exons, the duplication also predicts the existence of a novel peptide sequence within the protein which it is possible could confer altered biological properties on the mutant protein. Exon 1 codes for the 5' untranslated region (UTR) and a 42-amino acid sequence upstream of the first short consensus repeat (SCR) domain, ending in the first base of codon 43, a Guanine (G). The remainder of codon 43 (consisting of Guanine and Adenine (A)) is

encoded by exon 2, yielding Glycine (encoded by GGA) as the first amino acid of SCR 1. Exon 3 ends with the first base of codon 168, an Adenine. Splicing of another copy of exon 2 immediately downstream of exon 3 (as predicted by the genomic duplication in CFHR5 nephropathy) produces the codon AGA, predicting Arginine as the first amino acid of duplicated SCR 1 – see ***Figure 4.2-2***. The remaining residues of duplicated SCRs 1 and 2 (and the remaining SCRs 3-9) are identical to the wild type sequence. While Glycine is small, uncharged (at pH 7.4) and hydrophobic (with a hydropathy index of -0.4 (Kyte and Doolittle 1982)), Arginine is large, carries a positive charge and is the most hydrophilic amino acid (with a hydropathy index of -4.5). Although such a change could have functional consequences, possibly owing to altered conformation, if it occurs at or close to a functionally important site of a protein (indeed HIF2α erythrocytosis results from exactly this substitution in the region of a critical reactive proline residue, see Chapter 2), it seems somewhat less likely that this single amino acid substitution is responsible for altering function of the protein compared with the change in overall charge of the protein predicted by duplication of the whole of SCRs 1 and 2.

Figure 4.2-2 Amino acid sequence of first 6 exons of the translated mutant CFHR5
Note the duplicated exon 2' begins with Arginine (R, in red) rather than Glycine (G).

In addition to the electrostatic effects it is also possible that the critical functional change caused by the mutation arises from altered conformation of the protein. Perhaps the 2-SCR duplication or the glycine to arginine substitution reduces substrate accessibility to another part of the protein, impairing function. Conversely, it is possible that the mutant protein binds more strongly to an interacting partner owing to the presence of one or more binding sites within SCRs 1 and 2. Nonetheless, whatever the mechanism, this *in silico* prediction of the biochemical effect of the mutation was consistent with altered function of the protein, and hence provided a theoretical basis for CFHR5 nephropathy to result from this mutation.

4.2.2 Erythrocyte lysis

Since CFHR5 has biochemical properties which allow it to target to host surfaces (and is found co-localised with complement C3 in immune deposits *in vivo*) and the mutation could be predicted to impair this function, an experiment was performed which aimed to test the ability of mutant CFHR5 to localize with C3 on the surface of chick erythrocytes. Chick erythrocytes are spontaneously lysed by complement in human serum, resulting in C3 (and terminal pathway components) being deposited on the erythrocyte membrane fragments which can be isolated by centrifugation. Components of the serum which bind to C3 (and other deposited complement components) will therefore be depleted from the supernatant and enriched in the pellet. These components can be separated by electrophoresis and identified by immunoblotting. This experiment is summarized in ***Figure 4.2-3***, and the result demonstrated that, while the wild type CFHR5 protein is found predominantly in the pellet, the mutant $CFHR5^{12123\text{-}9}$ species was almost entirely found in the supernatant. This indicated that, in the relative proportions in which these proteins are found in the circulation of patients, the wild-type allele has a greater affinity for complement-lysed chick erythrocytes than does the mutant allele.

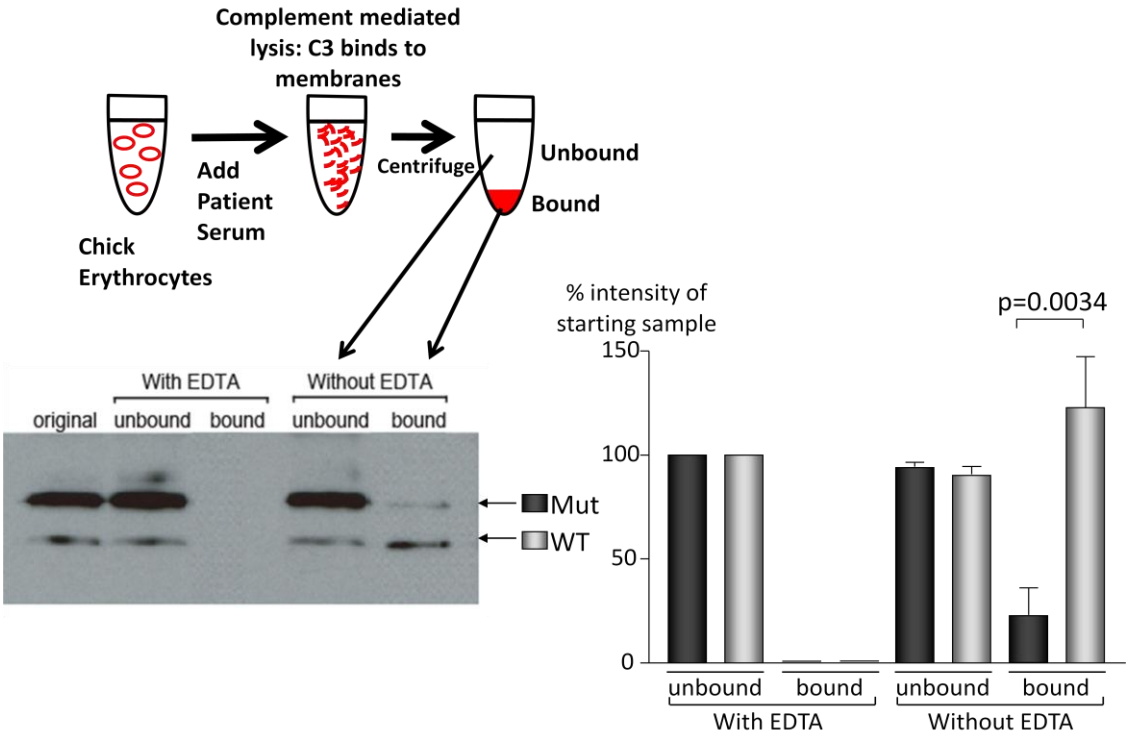


Figure 4.2-3 Chick erythrocyte lysis experiment

The mutant (Mut) CFHR5 protein is less abundant in the erythrocyte membrane pellet than it is in the supernatant. This contrasts with the wild type (WT) protein of which most is present in the pellet.

As a negative control, the experiment was performed in the presence of EDTA, which inhibits complement activation. In these conditions, there was complete absence of CFHR5 (wild type or mutant) from the pellet (which was composed of unlysed erythrocytes), providing evidence that non-specific binding of CFHR5 to the outside of the erythrocyte membranes was not occurring. In addition, immunoblotting confirmed the presence of C3 in the pellet (***Figure 4.2-4***).

Clearly, this experiment cannot determine which component(s) of the complement cascade CFHR5 is binding to, but it did provide the first direct evidence that the altered physico-chemical properties of mutant CFHR5 are sufficient to affect the ability of the protein to localize with complement to a surface.

I also observed that, when examining patient sera, the band corresponding to the mutant CFHR5 protein was of consistently greater intensity than that corresponding to the wild type CFHR5 protein. The antibody used for this analysis was monoclonal and raised against the whole protein. McRae et al demonstrated that this antibody recognized both the CFHR5^{1234} and CFHR5^{34567} peptides, suggesting that it recognized an epitope residing in SCR 3 or 4. This implies that higher affinity of the antibody for the mutant (compared with the wild type) protein is less likely to explain this discrepancy. One attractive possibility was that, in affected individuals, the wild type protein is preferentially bound to the glomerular electron dense deposits leaving relatively little of this species in the circulation. Conversely, since the mutant CFHR5 protein has lower affinity for surface-fixed complement, it would remain relatively abundant in the circulation, manifesting as more intense staining of the relevant band on Western blotting.

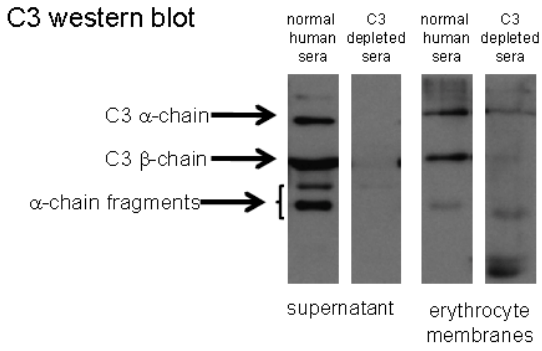


Figure 4.2-4 C3 is detectable in the lysed erythrocyte pellets

4.2.3 Heparin Binding – CFHR5 from serum

The observation that CFHR5 is able to bind to heparin (McRae, Duthy et al. 2005) raised the possibility that the mutant allele had altered ability to bind to heparin. In addition, heparin is among the most strongly negatively charged biological molecules – the change in charge predicted by the mutation (which is likely to render the mutant allele negatively charged at physiological pH) can be expected to result in an electrochemical force which opposes heparin binding. In order to test this, patient serum was passed through a heparin column (HiTrap, GE Healthcare Ltd, Buckinghamshire). The column was then washed and the bound components eluted with various concentration of NaCl. At each NaCl concentration, the eluate was blotted for CFHR5. While the wild type protein was only detected in the fraction eluted by 350 mM NaCl solution, the mutant band was more abundant in the fraction eluted at 300 mM (***Figure 4.2-5***), indicating that affinity for the heparin column of the mutant protein was reduced. In common with the erythrocyte lysis experiment, this experiment directly compared the affinity of the alleles to heparin in the relative concentrations in which they are present in the circulation of patients. The result supported the hypothesis that (and provided a potential mechanism by which) the mutation results in reduced ability of the protein to localise to host cells, which commonly express heparin-like moieties.

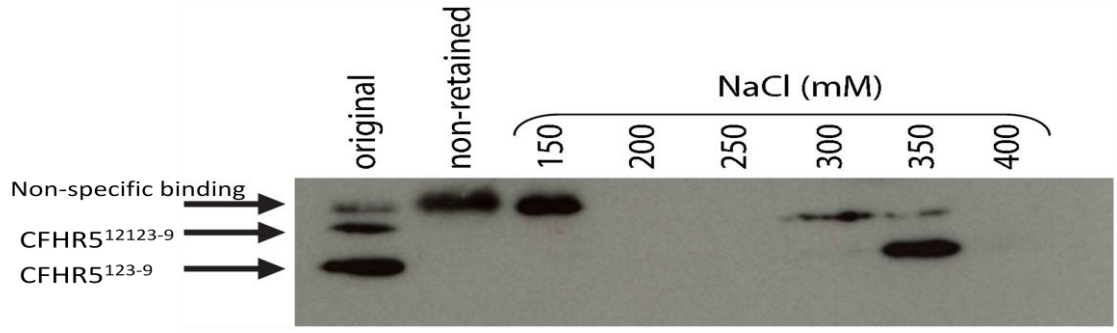


Figure 4.2-5 Elution of serum-derived CFHR5 from heparin columns
The mutant CFHR5$^{12123\text{-}9}$ is eluted at 300 mM NaCl whereas the wild type CFHR5$^{123\text{-}9}$ protein only becomes dissociated from the heparin column at 350 mM NaCl.

4.2.4 Generation of recombinant CFHR5$^{12123\text{-}9}$

While the assays of patient-derived serum indicated that the mutant protein had reduced affinity for complement-coated erythrocytes and for heparin, the possibility remained that a change other than the duplication of SCRs 1 and 2 was responsible for this altered behaviour – this possibility was unlikely because Western blotting allowed specific detection of the protein product of each allele and the behaviour of each was directly compared. Nonetheless, in order to understand better the functional effect of the mutation, and to exclude the possibility that the altered behaviour of the alleles was a consequence of an unidentified abnormality (presumably of a gene on the disease-linked haplotype) I performed *in vitro* mutagenesis of cloned cDNA of CFHR5 in order to allow expression of the recombinant alleles.

4.2.4.1 CFHR5 cDNA

CFHR5 cDNA was purchased from Open Biosystems (Thermo Scientific, Huntsville AL, USA) in the vector pCR-BluntII-TOPO which also contains a kanamycin resistance gene. Chemically competent DH5α *E. coli* were transformed with the plasmids, grown on agar plates impregnated with kanamycin and 5 colonies were picked for expansion. Plasmid DNA was extracted using a miniprep kit (QIAGEN) from aliquots of each culture and DNA from a single colony (which was subsequently used for all further experiments), was sequenced using primers complementary to the T7 (forward) M13 (reverse) promoter sequences upstream and downstream (respectively) of the cDNA for CFHR5. This confirmed firstly that the cDNA was identical to the database sequence for CFHR5 and secondly demonstrated the exact insertion site within the vector (***Figure 4.2-6***).

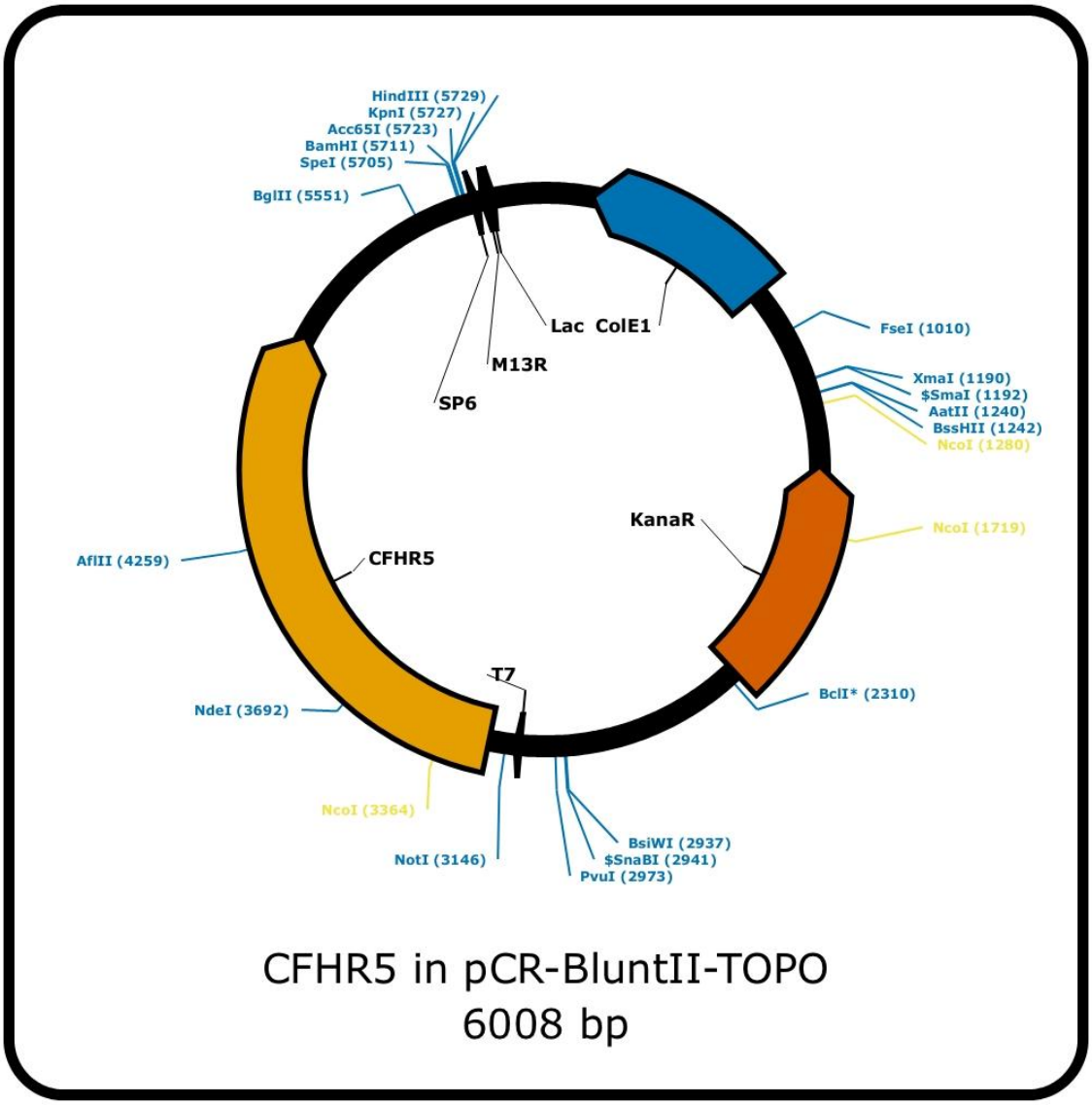


Figure 4.2-6 CFHR5 cDNA and vector
Showing unique restriction sites (blue) including the NdeI site and also the 3 NcoI sites (orange)

4.2.4.2 Amplification of the duplicated exons

Knowledge of the exact gene and vector sequence allowed identification of a unique restriction site (NdeI) which lay 15-bp from the 3' end of exon 3. I selected primers which would amplify exons 2 and 3 of the cDNA, but at the 5' end of the upstream primer (labelled "Exon_2_start") I inserted an oligonucleotide with the same sequence as the 25-bp at the 3' end of exon 3 (underlined in blue in ***Figure 4.2-7***).

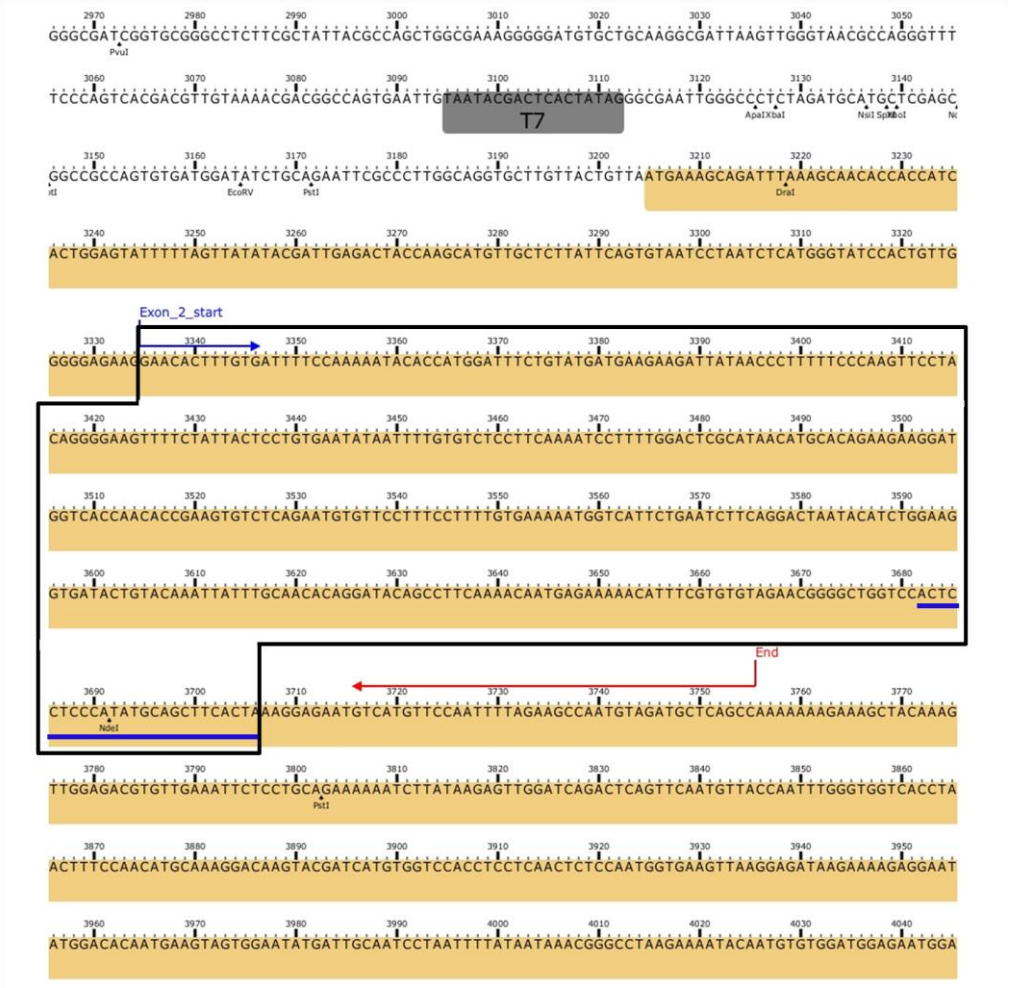


Figure 4.2-7 Plasmid sequence section
Showing: T7 promoter site (grey); CFHR5 coding sequence (orange); CFHR5 exons 2 and 3 (within the black box); and the sequences of the forward (blue arrow) and reverse (red arrow, 3'→5' complementary) are shown, as is the unique NdeI site.

This 5' overhang contained the NdeI site. Amplification of the CFHR5 plasmid with these primers was predicted to produce a 446-bp product containing the last 25-bp of exon 3, followed by the whole of exons 2 and 3 and the first 49-bp of exon 4 (***Figure 4.2-8***). Since there is an NdeI site 15-bp from the 3' end of exon 3, I subjected this PCR product to overnight digestion with this enzyme. This was predicted to remove the 5' 10-bp and the 3' 64-bp, producing a 372-bp product consisting of the last (i.e. 3') 15-bp of exon 3, followed by the whole of exon 2, followed by all but the last 15-bp of exon 3, with (identical) "sticky" ends suitable for ligation into an NdeI-digested template. I next digested the pCR-BluntII-TOPO-

CFHR5 plasmid with NdeI overnight, and treated the product with alkaline phosphatase which results in loss of the 5' terminal phosphate moieties of the linearised plasmid, preventing their self ligation in the presence of DNA ligase. Both digests were electrophoresed on a 2% agarose gel and the bands were purified (QIAGEN).

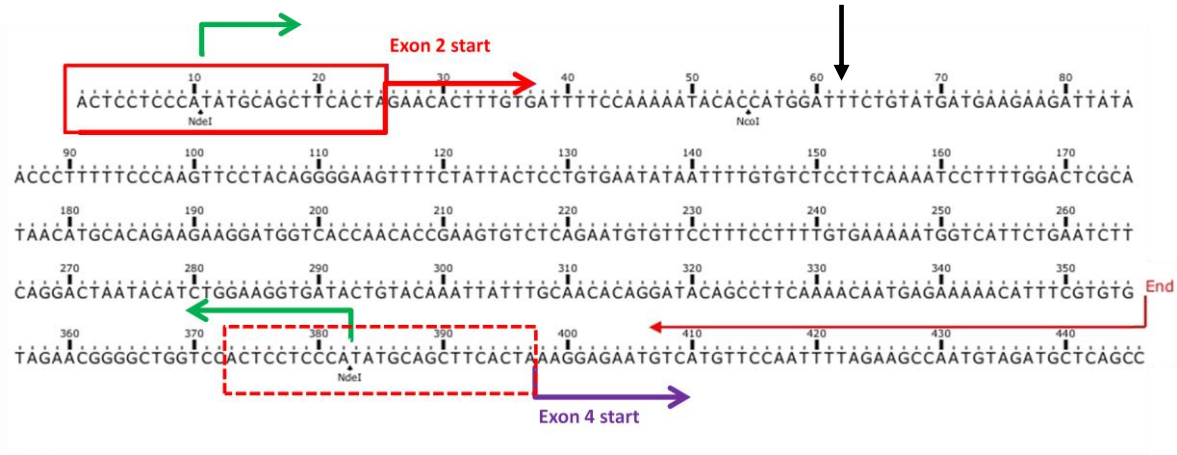


Figure 4.2-8 Plasmid PCR-amplicon

Amplicon resulting from amplification of CFHR5 cDNA using forward primer consisting of the 3' 25-bp of exon 3 (red box) placed 5' to the first 12-bp of exon 2 (forward arrow) and a reverse primer which was the reverse-complement of an oligonucleotide present in exon 4 (purple arrow). Digestion of this amplicon with NdeI produced a 372-bp fragment (between green arrows) consisting of the last 15-bp of exon 3 followed by all of exon 2 followed by all but the last 15-bp of exon 3. Note the NcoI site 44-bp from the 5' end (and 328-bp from the 3' end) of this fragment (black arrow).

4.2.4.3 Ligation of amplified exons 2 and 3 into gene

Since a single restriction enzyme had been used both to generate sticky ends on the insert and to linearise the plasmid, a number of possible outcomes could be anticipated to result from ligation. Firstly, self ligation of the linearised plasmid to form the wild-type CFHR5 in the vector; secondly, correct insertion of the digested PCR product to form the desired mutant in the vector; thirdly, incorrect insertion of the digested PCR product inverted with respect to the plasmid; fourthly, insertion of multiple copies of the PCR product into the plasmid. I considered that the first possibility was unlikely to occur owing to the prior treatment of the digested plasmid with alkaline phosphatase – lack of the 5' phosphate group on both the sense and antisense strands would prevent formation of a covalent bond between the hydroxyl group on

the 3’ end and the 5’ end of the linearised plasmid. I predicted that ligation of the insert would occur in the desired and inverted orientation with equal likelihood, but that insertion of multiple copies of the insert would be unlikely unless the insert was in significant excess with respect to the vector.

I therefore performed the ligation reactions and PCR product:plasmid molar ratios of 3:1 and 10:1. The ligated products were then used to transform chemically competent DH5α *E. coli* which were then plated onto kanamycin-impregnated agar plates. 10 individual colonies were picked for each molar ratio, allowed to replicate and plasmid DNA was extracted from an aliquot of each one.

4.2.4.4 Diagnostic digestion of mutant construct

Diagnostic digestion with NcoI was predicted to produce 4 bands of size 372, 439, 1645 and 3924-bp if the insert was correctly inserted (with 372-bp band being lost in the absence of an insert). Inversion of the insert was predicted to produce bands of sizes 439, 654, 1645 and 3634-bp respectively (***Figure 4.2-9***).

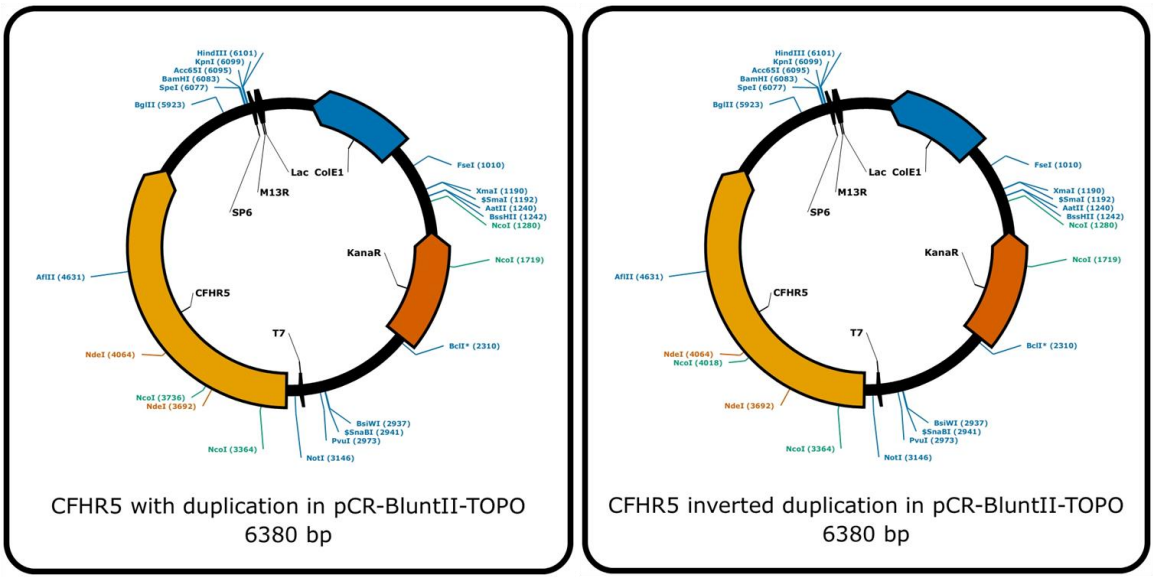


Figure 4.2-9 The desired and inverted ligation products

The desired product contains NcoI restriction sites at positions 1280, 1719, 3364 and 3736, predicting bands of size 372, 439, 1645 and 3924-bp following NcoI digestion. The inverted product contains NcoI sites in positions 1280, 1719, 3364 and 4064, predicting bands of 439, 654, 1645 and 3634-bp respectively.

This was performed and confirmed the presence of the predicted bands in none of the 10 colonies transformed with product of ligation at a 3:1 molar ratio, but 3 of the 10 colonies transformed with 10:1 molar ratio ligation product (see ***Figure 4.2-10***). In the remaining 7 colonies, the 372-bp band was absent – consistent with failure to ligate the PCR product into the vector. One colony with the desired diagnostic digest was selected and the gene sequenced (again using T7 and reverse M13 primers), confirming that the desired recombinant mutant allele had been generated.

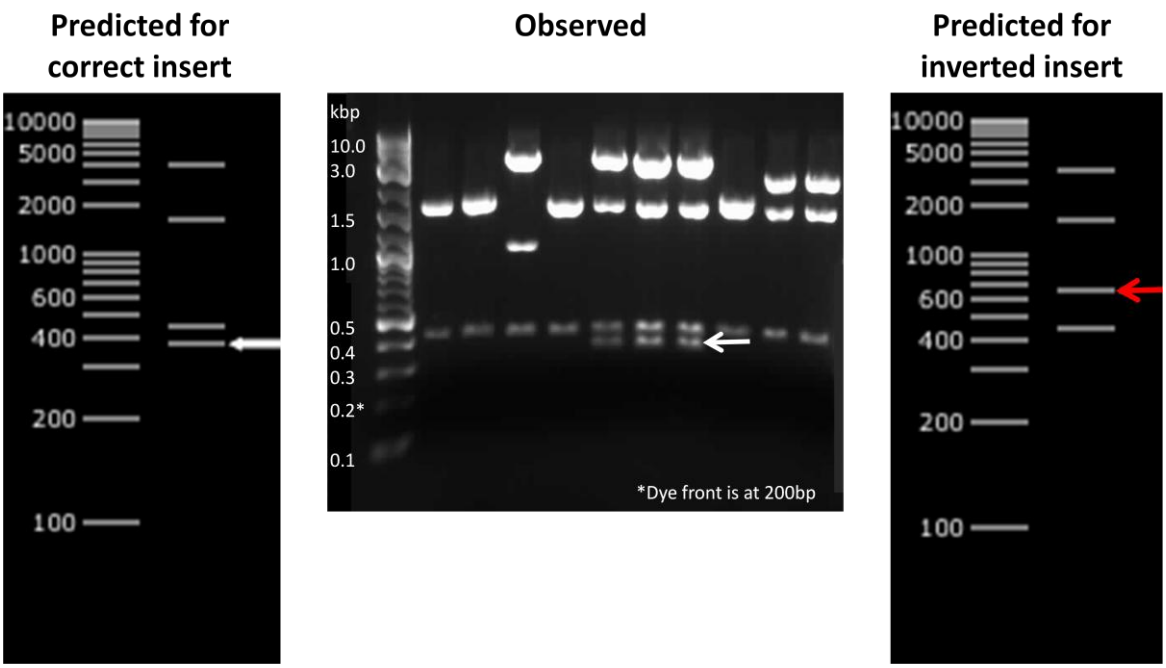


Figure 4.2-10 Predicted and observed gels
The 372-bp fragment (white arrow) is only predicted to be present when the insert is present in the correct orientation. Inversion of the insert is predicted to produce an aberrant band of 654-bp (red arrow).

The subcloning, heparin and glomerular binding and cofactor assays were performed by Drs Matthew Pickering and Elena Goicoechea de Jorge. The experiments are included in this thesis because I was involved in the experimental design and interpretation, and because the insights they give into the functional effects of the CFHR5 mutation are important considerations for material presented later.

4.2.4.5 Sub-cloning into pCI-Neo and expression in Cos7 cells

In order to allow efficient expression of the mutant protein in mammalian cells expressing the SV-40 large T antigen, the recombinant CFHR5 alleles which I had made were cloned into pCI-Neo vector (obtained from Promega) which contains an SV40 enhancer/promoter. This was achieved by amplification of the wild type and mutant genes in pCR-BluntII-TOPO using primers incorporating (in their 5' ends) a SalI restriction site (5' end of the gene) and a NotI restriction site (3' end of the gene – i.e. the reverse primer). Amplification was performed using Platinum Taq High Fidelity (Invitrogen), which has proof-reading ability via its 3' to 5' exonuclease activity. The PCR products were digested with SalI and NotI to produce sticky ends and the resulting digested amplicons ligated into the pCI-Neo vector which had been cut with SalI and NotI to produce the products shown in ***Figure 4.2-11***. These constructs were used to transform competent *E. coli* bacteria and amplified in culture. The plasmid DNA was extracted and the inserts sequenced using T7 (forward) and T3 (reverse) primers. This confirmed correct insertion of the alleles into the vector and excluded the presence of substitutions, insertions or deletions which might have occurred during the PCR amplification procedure.

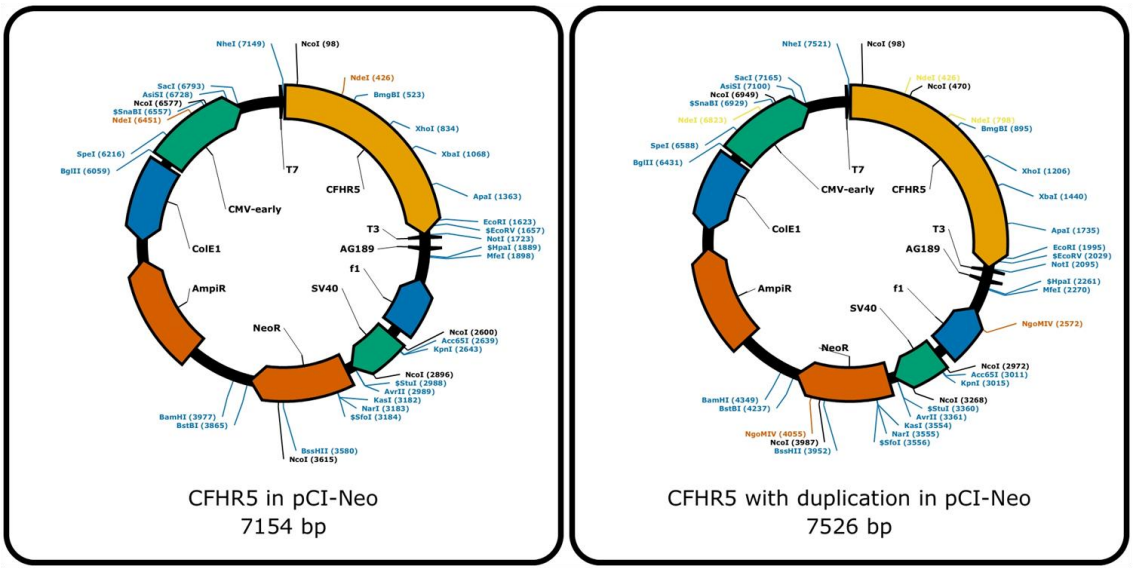


Figure 4.2-11 CFHR5 in expression vector pCI-Neo

pCI-Neo-CFHR5 wild type and mutant plasmid DNA was transiently transfected into Cos7 cells using calcium chloride and grown under serum-free media. After incubation for 48 hours, aliquots of the media were blotted for CFHR5 and this demonstrated the presence of immunoreactive bands corresponding to the size predicted for $CFHR5^{123\text{-}9}$ and $CFHR5^{12123\text{-}9}$ respectively (***Figure 4.2-12***). Next, an enzyme-linked immunosorbent assay was performed using rabbit polyclonal and mouse monoclonal antibodies. This demonstrated similar levels of expression of both alleles in the supernatant, consistent with the result from Western blotting (***Figure 4.2-13***).

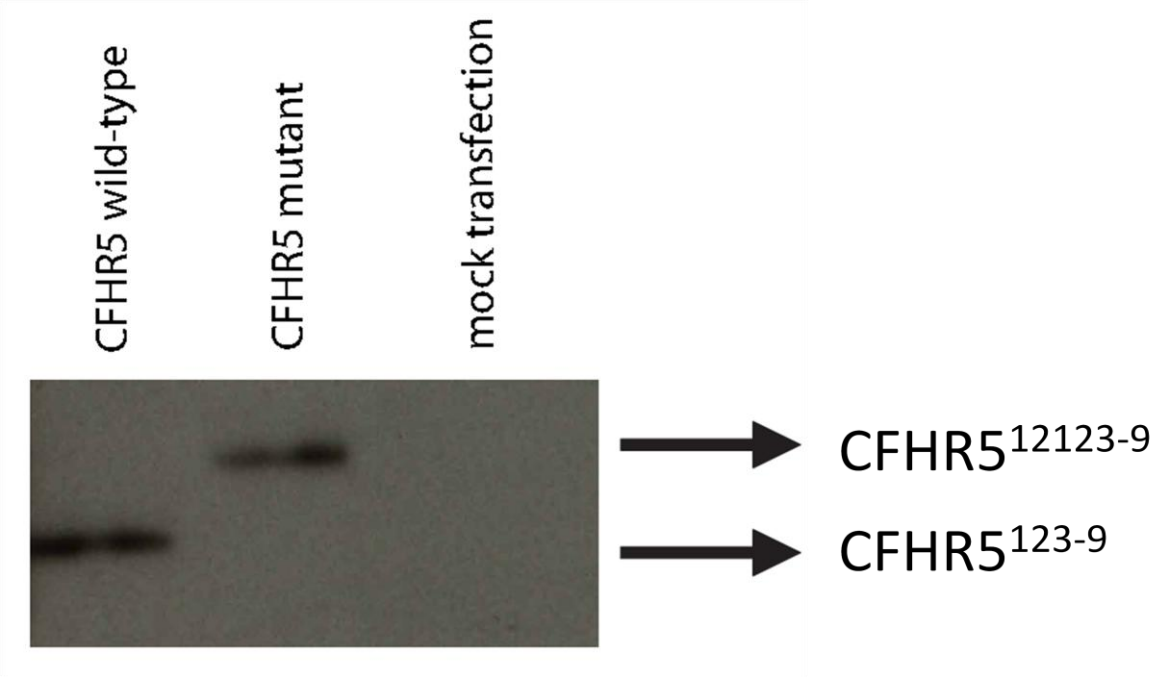


Figure 4.2-12 Expression of CFHR5 wild type and mutant proteins in Cos7 cells

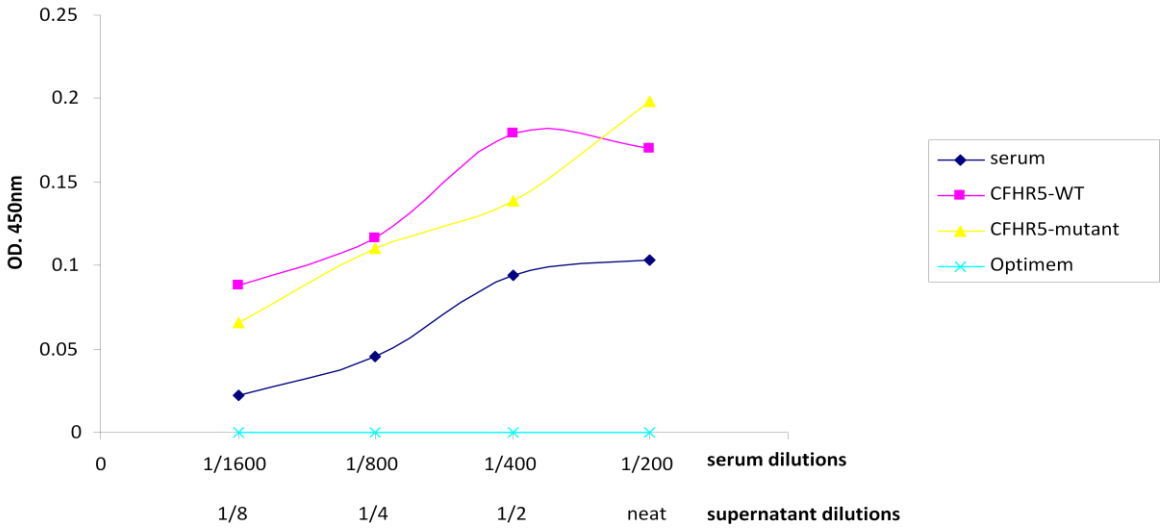


Figure 4.2-13 Enzyme-linked immunosorbent assay for CFHR5

ELISA was performed on serum and culture medium from Cos7 cells transfected with wild type or mutant CFHR5

4.2.5 Heparin-binding – recombinant CFHR5

Culture media containing each protein was passed through HiTrap heparin columns and eluted at increasing NaCl concentration (***Figure 4.2-14***). While the wild type CFHR5 was predominantly eluted at 350mM NaCl, the CFHR5$^{12123-9}$ protein was detectable in the eluate of 250 mM NaCl and was predominantly released in the 300 mM fraction. This protein was not detected in the 350 mM fraction.

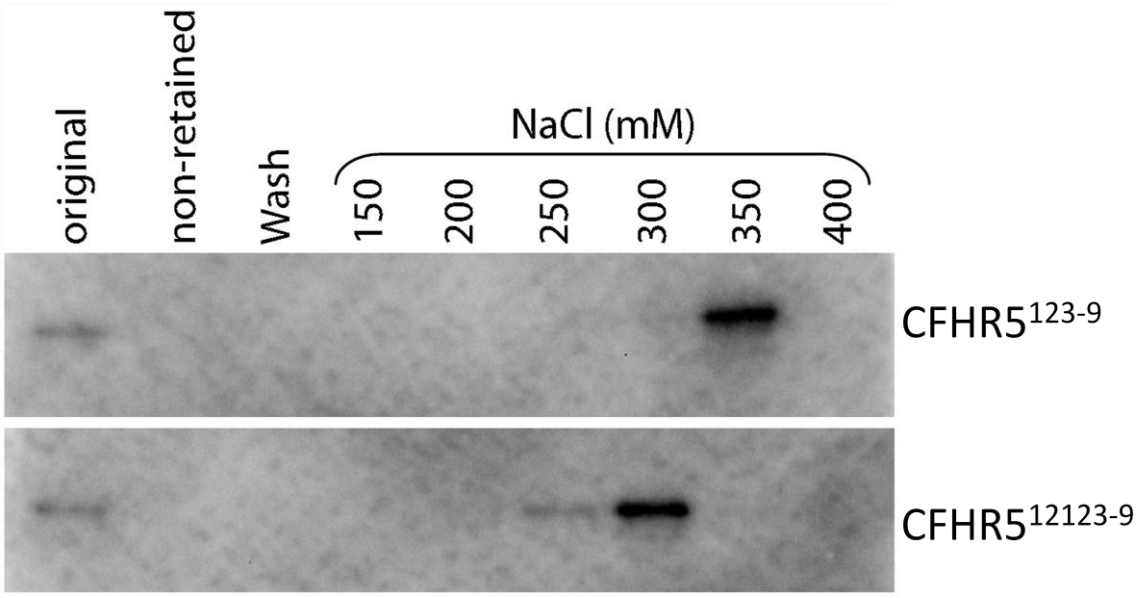


Figure 4.2-14 Elution of recombinant CFHR5 proteins from heparin columns
Recombinant CFHR5 proteins were allowed to bind to heparin columns and eluted with increasing NaCl concentrations. The mutant protein is eluted at a lower concentration than the wild type, consistent with lower affinity for the column.

We concluded that the mutation causes CFHR5 to have reduced affinity for heparin.

4.2.6 Glomerular binding – recombinant CFHR5

Since CFHR5 is present in the glomerular deposits of C3 (McRae, Cowan et al. 2001) and the theoretical and heparin binding data were consistent with impaired ability of the mutant protein to bind to anionic surfaces, an experiment was performed which compared the ability

of the recombinant proteins to bind to glomerular sections coated with C3. The substrate used for this experiment was glomerular sections from mice deficient in CFH (*cfh-/-*). These animals spontaneously develop C3 accumulation along the glomerular basement membrane (Pickering, Cook et al. 2002). The sections were first incubated with 5% normal human serum, washed and then incubated with mouse monoclonal anti-CFHR5 antibody. The sections were then incubated with fluoroscein isothiocyanate-labelled goat anti-human IgG antibodies in order to visualise and quantify any deposition of CFHR5 in the glomeruli. For comparison of the distribution, adjacent sections (from the same animal) were stained with fluorescein isothiocyanate (FITC)-labelled mouse anti-C3 antibodies, and as a negative control, sections from mice deficient in both *cfh* and *c3* were stained for CFHR5. This revealed that CFHR5 from human serum bound to the mouse glomeruli in a similar distribution to C3, but did not bind in the absence of C3 (***Figure 4.2-15***). This result was expected in view of the previous observation that CFHR5 co-localizes with C3 in human glomeruli (Murphy, Georgiou et al. 2002) and provided proof of concept that the mouse complement system could act as a model for (at least some of) the biochemical functions of human CFHR5.

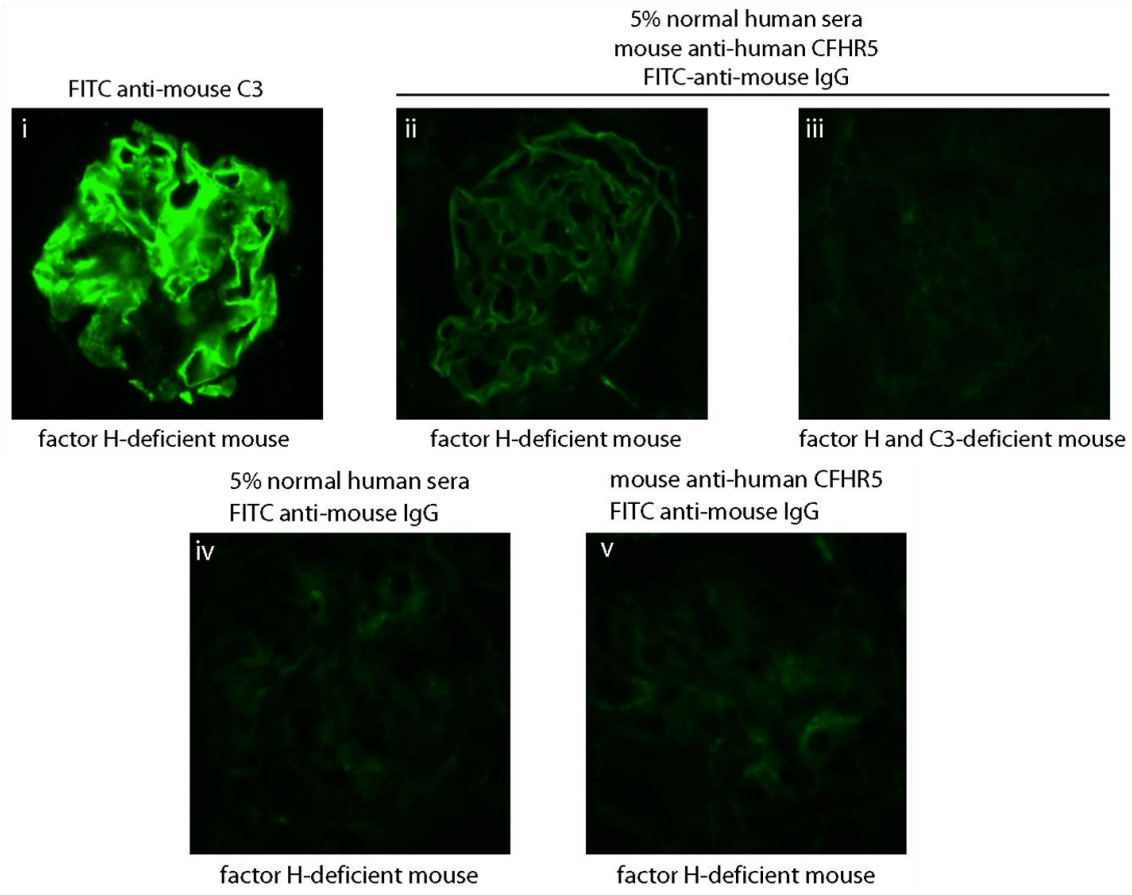


Figure 4.2-15 CFHR5 in complement-coated mouse glomerular sections

Human CFHR5 is detected in CFH-deficient mouse glomerular sections following incubation with normal human serum in a similar distribution to C3. Human CFHR5 is not detected when the mouse is deficient for both CFH and C3, or if the glomerular sections are not incubated with human sera.

The erythrocyte lysis and heparin binding assays were both consistent with reduction in affinity of the mutant protein for negatively charged host surfaces and it was therefore predicted that the mutant protein would bind less well to the mouse glomeruli. Serial dilutions of supernatant were tested and the supernatant containing the mutant protein resulted in significantly lower fluorescence intensity compared with that containing the wild type protein (***Figure 4.2-16***).

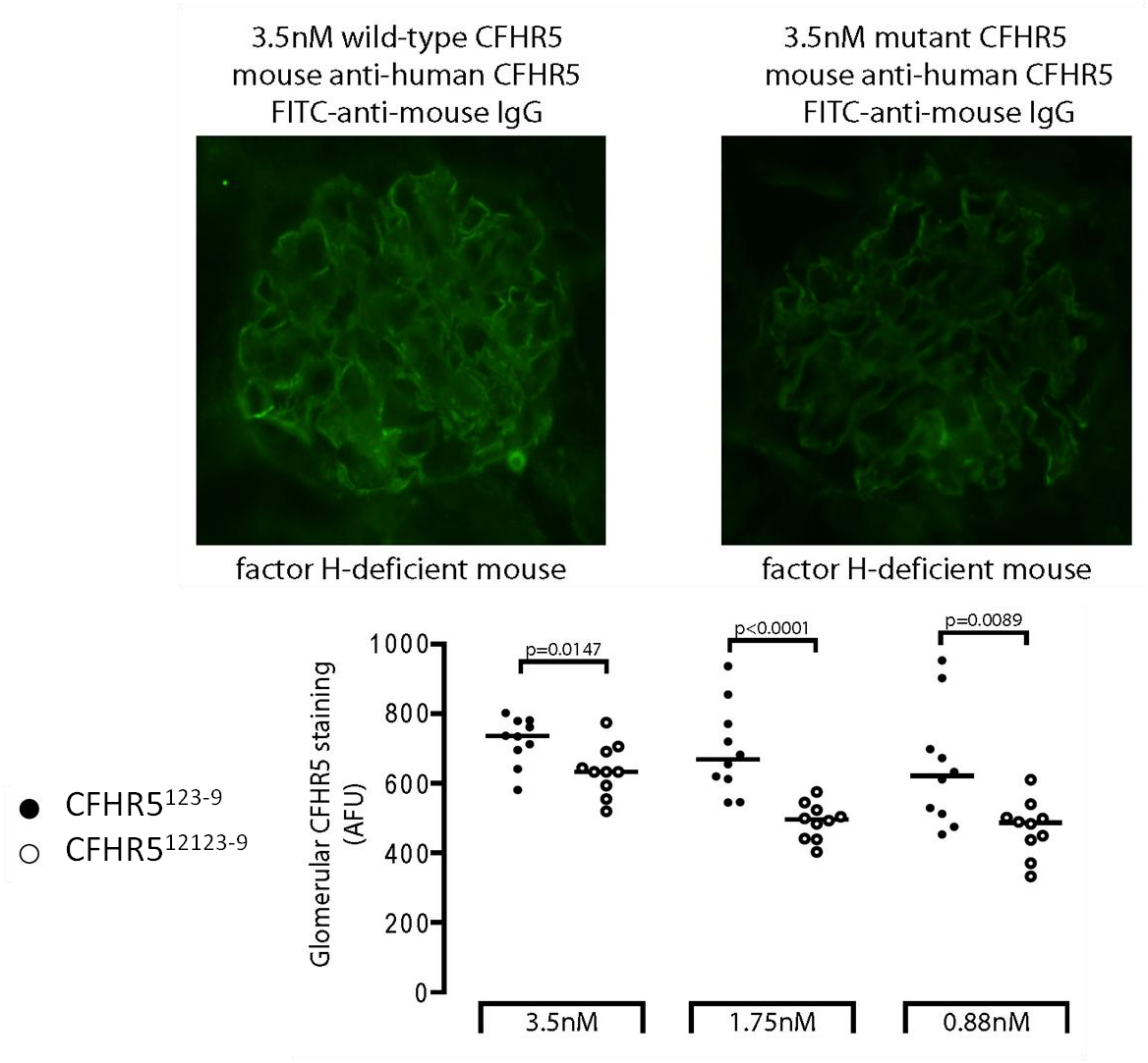


Figure 4.2-16 Impaired binding of mutant CFHR5 to complement-coated glomeruli

Significantly less fluorescence was detected when recombinant mutant CFHR5 was incubated with CFH deficient mouse glomerular sections compared with the wild type, at various concentrations of recombinant protein.

4.2.7 Cofactor activity – recombinant CFHR5

The functional data reported above were consistent with a reduced ability of the mutant protein to localize to host surfaces but did not address the question of whether the mutant protein had altered ability to regulate C3 turnover, in addition to the altered ability to localize to surfaces.

Since CFHR5 had previously been shown to have the ability to act as a cofactor for Factor I-mediated cleavage of C3b to iC3b, the ability of the mutant protein to perform this function was assessed. C3b (Calbiochem, CA, USA) was generated from C3 by partial digestion with trypsin. 50 ng of C3b was added to 10 ng Factor I (Calbiochem, CA, USA) and serial dilutions of wild type or mutant CFHR5 were added to a total volume of 16.5 µL in 20 mM phosphate buffer (pH 6.0). The mixture was incubated overnight at 37°C, the reaction then stopped with 5 µL SDS sample buffer and the solution blotted for C3 (using goat anti-C3c polyclonal antibodies). Surprisingly, while no cleavage of the C3α chain was detected in solution containing 7 nM or less wild type CFHR5, cleavage was detected with as little as 3.5 nM CFHR5$^{12123\text{-}9}$ and no intact C3α chain was detected in the presence of 7 nM CFHR5$^{12123\text{-}9}$. A representative blot is shown in ***Figure 4.2-17***. This experiment indicated that, rather than coding for a simple loss-of-function variant of CFHR5, duplication of exons 2 and 3 produces a protein which has enhanced cofactor activity (at least in this assay).

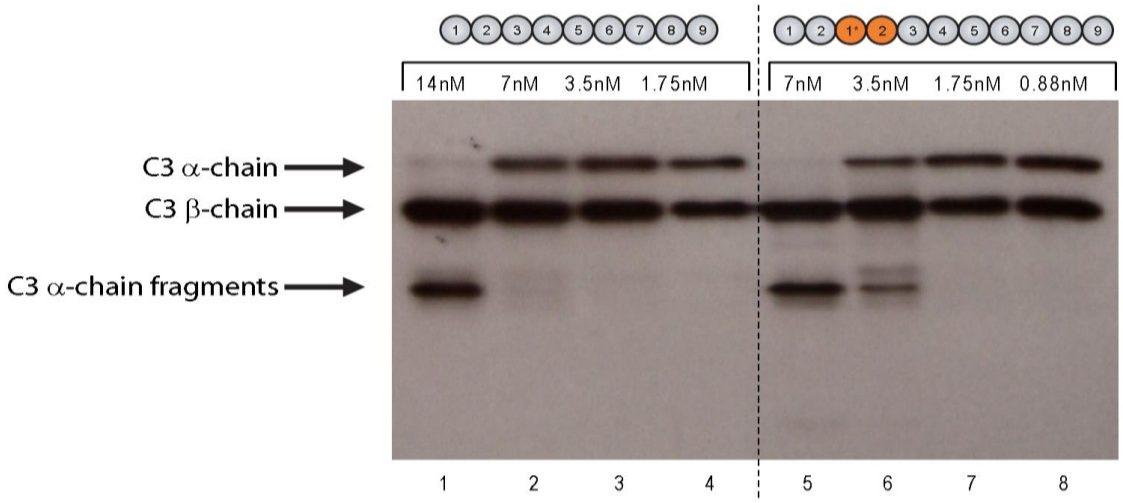


Figure 4.2-17 Recombinant CFHR5 cofactor activity

Cleavage of C3α in the presence of complement factor I was detected with lower concentrations of the mutant CFHR5 protein (lanes 5 to 8) compared with the wild type protein (lanes 1 to 4)

4.2.8 Insights from clinical observations

While experiments using CFHR5 and CFHR5$^{12123\text{-}9}$ isolated from serum or generated from recombinant DNA can provide some evidence of the biochemical effects of the mutation, two clinical observations have provided particular insight into the pathophysiology of the disease and have implications for its treatment.

4.2.8.1 Transplant recurrence

To date, 3 individuals with CFHR5 nephropathy have undergone renal transplantation. The first was individual III-1 from family 1 (see Chapter 3). This individual was born in Cyprus and was an obligate carrier of the CFHR5 duplication. A native renal biopsy performed in his fifth decade revealed C3GN and he reached end stage renal failure and received a cadaveric renal allograft at the age of 49. He died 11 years later from a myocardial infarction at which time his serum creatinine was 89 µmol/l. The allograft was never biopsied. The second individual (III-2 from family 3, see ***Figure 5.2-4***, Chapter 5) resided in Cyprus and had presented at end stage renal disease and hence had never been biopsied. Molecular analysis (performed on stored genomic DNA *post mortem*) confirmed the presence of duplication of exons 2 and 3 of CFHR5 in him. This individual had undergone cadaveric renal transplantation but lost the (well-functioning) graft and his right leg 10 years later to cholesterol embolic disease following coronary angiography. He died 6 years after that. Both of these individuals therefore had no clinical or histological evidence of disease recurrence following transplantation.

The third individual was a Cypriot man (designated 115301, see Chapter 5) with end stage renal failure secondary to CFHR5 nephropathy (diagnosed by the presence of synpharyngitic macroscopic haematuria, C3GN on native renal biopsy aged 34 and heterozygosity for the exon 2/3 duplication of CFHR5) who underwent cadaveric renal transplantation at the age of 53 with a kidney from a 66-year-old donor. There were no surgical complications and immunosuppression was with alemtuzumab and corticosteroids in the perioperative period,

and tacrolimus which was continued after discharge on postoperative day 8. Serum creatinine stabilised at 186 µmol/l and there was persistent microscopic haematuria with a single episode of macroscopic haematuria which was not related to instrumentation of the urinary tract. Allograft biopsy performed on postoperative day 46 demonstrated subendothelial and mesangial electron dense deposits (***Figure 4.2-18***). There was C3GN, with isolated granular staining for C3, but no detectable deposition of immunoglobulins. In this biopsy, there was no mesangiocapillary glomerulonephritis, and the impression of the histopathologist (Professor Cook) was that the suboptimal allograft function could not be explained by the glomerular changes.

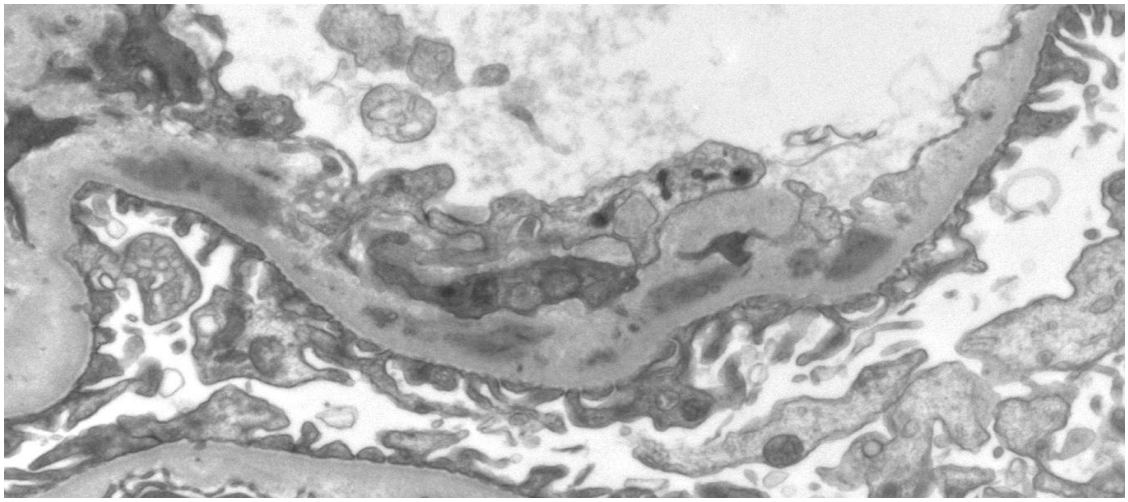


Figure 4.2-18 Recurrent disease following transplantation
Kidney allograft biopsy performed in an individual with CFHR5 nephropathy 46 days after transplantation showing subendothelial electron dense deposits.

Whatever the direct implications for this patient, this observation demonstrates that C3GN in CFHR5 nephropathy can recur in an allograft and proves that the disease results from an abnormality of a circulating factor, as opposed to a defect of local complement regulators synthesised within the kidney.

The systemic nature of the disease has two major clinical implications: firstly that organs other than the kidneys might be susceptible to the disease; and secondly that therapeutic

interventions which modulate circulating CFHR5 levels may be of benefit in CFHR5 nephropathy.

4.2.8.2 Therapy in CFHR5 nephropathy: anecdotal experience

An individual (designated 116301, see Chapter 5) had been diagnosed with MCGN following a renal biopsy performed to investigate macroscopic haematuria at the age of 18 years. Subsequent review was consistent with C3GN (i.e. there was immunostaining for C3 but not immunoglobulins and no dense transformation of the basement membrane) and molecular testing confirmed that the underlying disease was CFHR5 nephropathy. Ten years later there was the first evidence of renal dysfunction and proteinuria and the biopsy was repeated – again revealing C3GN but now with some tubular and mesangial scarring. He remained well until the age of 32, at which time there was a further rise in serum creatinine to approximately 200 µmol/l. Treatment with mycophenolate mofetil was commenced, but over the following 6 months the renal function deteriorated progressively and renal replacement therapy was commenced.

A 44-year-old individual (III-2 from family 2, for clinical history see Chapter 3) exhibited a rising creatinine during the first half of 2009, from a baseline of approximately 180 µmol/l in 2008. He was diagnosed with a urinary tract infection and subsequently underwent a bladder neck incision operation in March 2009. Following this procedure, the creatinine progressively rose to approximately 230 µmol/l and he was administered 1 g of cyclophosphamide intravenously at the start of June, with a further 1 g one month later. Immediately after the first dose was administered there was macroscopic haematuria and blood tests revealed a rising creatinine (see ***Figure 4.2-19***). The macroscopic haematuria became more severe after the second dose of cyclophosphamide and no further doses were given. The creatinine continued to rise, reaching 270 µmol/l prior to the second dose, with a continued gradual rise over the next 2 months. During this period the patient experienced persistent macroscopic

haematuria, nausea, weight loss and poor appetite. In September 2009 he was admitted to hospital with non-specific symptoms (feeling 'very unwell' with a presumed bacterial infection) and an acute deterioration in renal function, with a serum creatinine which peaked at approximately 620 µmol/l. Following treatment with intravenous rehydration and antibiotics the creatinine fell and at discharge it was 375 µmol/l. Over the ensuing 8 weeks, the creatinine fell slowly, however the macroscopic haematuria persisted.

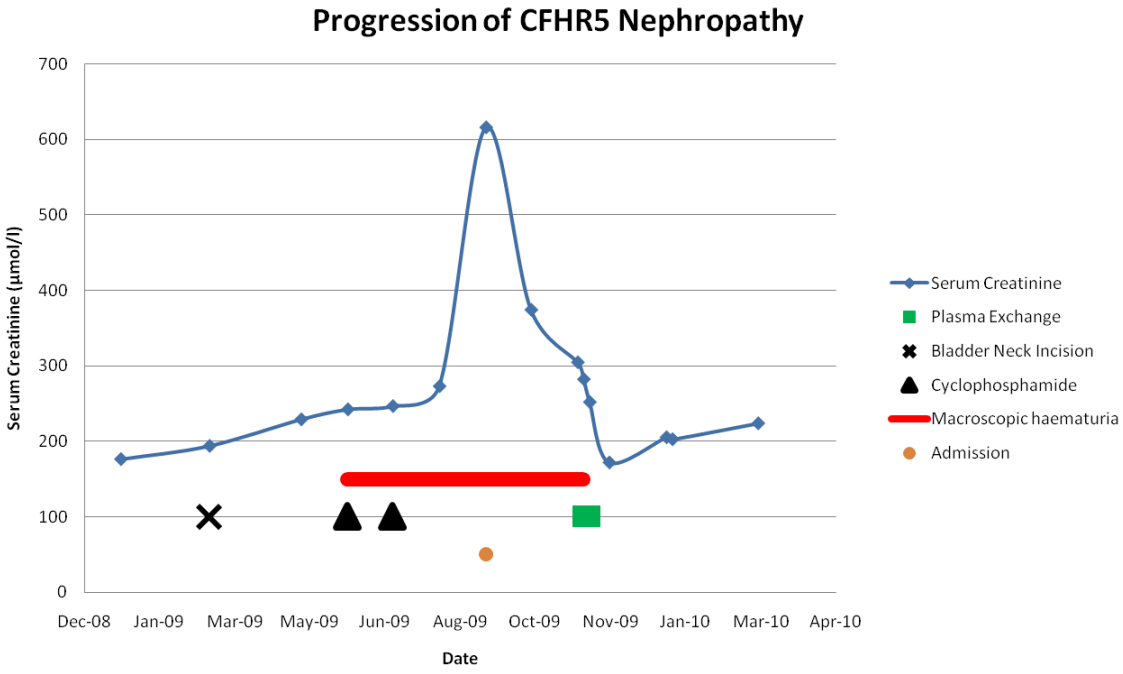


Figure 4.2-19 Treatment of CFHR5 nephropathy

Rising serum creatinine in an individual with CFHR5 nephropathy in the context of macroscopic haematuria. There was no response to immunosuppression with corticosteroids and cyclophosphamide, and admission with a presumed bacterial infection was associated with sudden deterioration in renal function. Macroscopic haematuria ceased and recovery of function to baseline was seen immediately following plasma exchange.

In view of the known abnormal circulating protein (i.e. CFHR5$^{12123\text{-}9}$) and the recently observed recurrence of the disease in a transplant, I considered that normalising the circulating concentrations of complement proteins using plasma exchange would have a good chance of therapeutic benefit, similar to other genetic diseases of complement regulation (Noris, Bucchioni et al. 2005). Because it was not possible to be certain whether supplementation of

wild type CFHR5 or removal of mutant CFHR5 would be needed to provide the maximum chance of benefit, in consultation with the clinician looking after him (Dr Pierides), the patient underwent 4 cycles of plasma exchange, each against 1.5 l fresh frozen donor plasma and 1.5 l 5% human albumin solution. On completion of this treatment there was a very marked clinical improvement with cessation of the macroscopic haematuria and an abrupt fall in the serum creatinine to under 180 µmol/l. Over the following 6 months, the serum creatinine has gradually risen, with a most recent reading of 223 µmol/l.

Western blotting of serum samples before and after each exchange revealed reduction in the levels of the mutant protein, more apparent after the third and fourth exchanges. There was no obvious change in the intensity of the band corresponding to the wild type protein (***Figure 4.2-20***).

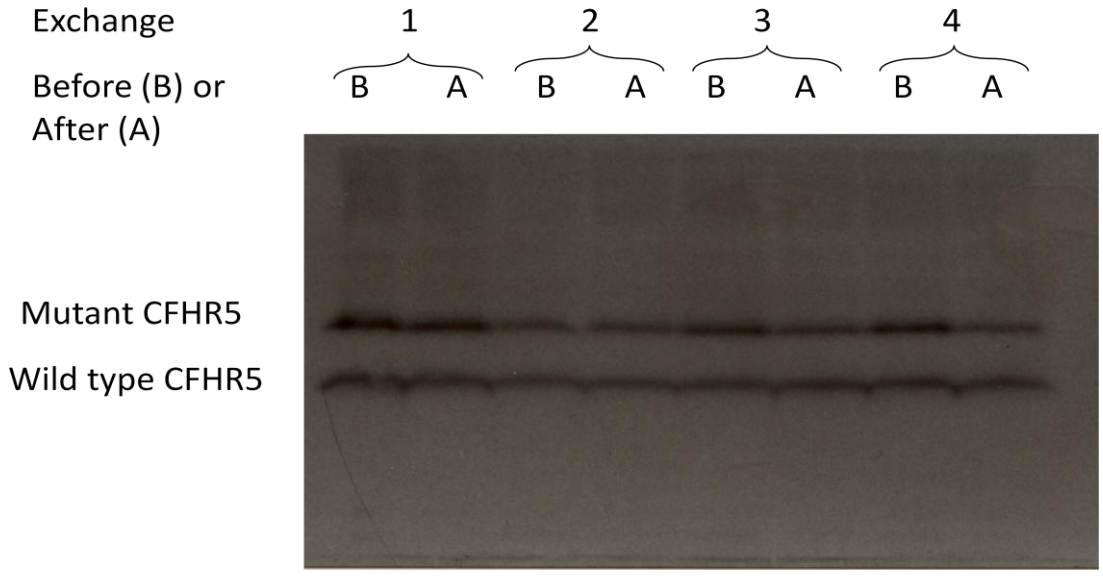


Figure 4.2-20 CFHR5 protein before and after plasma exchange

Western blot of serum for CFHR5 immediately before and after treatment with plasma exchange on 4 consecutive days. There is a detectable reduction in the amount of the mutant protein following exchanges, with apparent recovery to pre-treatment levels within 24 hours.

4.3 Discussion

The full range of biochemical properties, interactions and physiological functions of CFHR5 are incompletely understood. In this context, a complete appreciation of the biological effects of the observed duplication of 2 SCR domains is impossible. However, it is clear that previously reported biochemical properties of CFHR5 are altered by the presence of the mutation: $CFHR5^{12123\text{-}9}$ displays reduced affinity for negatively charged molecules and surfaces (namely heparin and complement-coated eukaryotic surfaces); appears to be more abundant than the wild-type CFHR5 protein in the circulation of patients; and has increased ability to act as a cofactor for factor I-mediated cleavage of C3b.

It is worth noting that, even though the methods of detection and quantification of the CFHR5 alleles (i.e. ELISA and Western blotting) relied on antibody binding, a significant difference in the affinity of the antibody for the protein products of the two alleles would not explain the differences observed: for both the erythrocyte lysis and heparin binding experiments, the quantity of each protein in different fractions of the same experiment was compared (i.e. pellet versus supernatant in the erythrocyte assay and each eluate in the heparin binding assay) and this ratio was compared between the two proteins.

If the explanation of the reduced glomerular staining for $CFHR5^{12123\text{-}9}$ in the mouse kidney sections treated with recombinant CFHR5 alleles was reduced affinity for the antibody (rather than reduced affinity for the glomerular deposits) then the quantity of $CFHR5^{12123\text{-}9}$ administered to the sections would have been similarly underestimated, which would tend to oppose this error. A similar argument can be deployed in the case of a hypothetically increased affinity of the antibody for the mutant protein.

Finally, in the original publication describing the antibody, it was shown to detect both $CFHR5^{1234}$ and $CFHR5^{34567}$ peptides (McRae, Duthy et al. 2005), implying that it recognised epitope(s) other than those found in SCRs 1 and 2. I therefore consider it impossible for *all* the

observed differences between the two alleles in the binding assays to be an artefact of a difference in antibody affinity, and unlikely for *any* of them to have resulted from this type of artefact.

4.3.1 CFHR5$^{12123\text{-}9}$: A gain-of-function mutation?

It is not immediately obvious how *increased* cofactor activity (leading to enhanced ability to degrade C3b) could be the mechanism by which CFHR5$^{12123\text{-}9}$ is responsible for CFHR5 nephropathy. One possibility is that this observation resulted from a flaw in the design of the experiment: since the read-out for the cofactor reaction was degradation of C3b (rather than quantification of CFHR5), interpretation of the assay is highly dependent on the accuracy of quantification of CFHR5 administered. Nonetheless, it seems unlikely (given, for instance, the strong staining for the mutant band in patient sera) that the concentration of recombinant mutant protein used in the experiment was inadvertently so much higher (requiring a *reduced* affinity for the antibody) than that of the wild type band – especially in view of the greater than 2-fold increase in concentration of the wild type allele required for the reaction to proceed. It therefore seems more likely that the observed enhanced cofactor activity reflects a genuine difference in the properties of the proteins.

While iC3b (the product of factor I-mediated cleavage of C3b) does have pro-inflammatory effects (via its ability to act as a ligand for complement receptors on phagocytic cells), it lacks C3 convertase activity (see Chapter 3). Therefore, an abnormality which resulted (purely) in increased cleavage of C3b to form iC3b would not result in increased C3 turnover – in fact accelerated clearance of C3b would tend to *reduce* alternative pathway activation by reducing the levels of C3 convertase. Importantly, evidence from both humans and mice with genetic complement abnormalities has shown that CFI activity (and presumably therefore iC3b generation) is required for complement-mediated renal damage. As discussed in Chapter 3, humans with homozygous deficiency in CFI, while manifesting C3 deficiency (with what is

detected predominantly present as C3b, presumably owing to the uncontrolled conversion of C3 to C3b via the alternative pathway), do not always exhibit glomerulonephritis and in particular, no cases of C3GN or DDD have been reported in these individuals (Rose, Paixao-Cavalcante et al. 2008; Licht and Fremeaux-Bacchi 2009). In addition, mice deficient for both *cfi* and *cfh,* do not exhibit glomerular damage until after the administration of exogenous source of cfi which allows iC3b to be generated (Rose, Paixao-Cavalcante et al. 2008). Together these data are most likely to be explained by a pathogenic role of iC3b in mediating glomerular damage. Although it is true that circulating C3 levels were normal in all affected individuals studied, compatible with the absence of generalised C3 consumption, the hypothesis that enhanced C3b cleavage *alone* (i.e. without autocatalytic upregulation of C3b production by alternative pathway dysregulation) is sufficient to cause glomerulonephritis owing purely to increased iC3b production is intriguing but intuitively unlikely: iC3b may be necessary but has never been shown to be sufficient to cause C3-mediated glomerular damage in the absence of excessive C3b generation. Nonetheless, the fact that the evolution of the complement system has resulted in the requirement of CFI for a cofactor in order to cleave C3b suggests that regulation of this step is of critical importance in the pathway, implying that abnormal kinetics of this reaction could result in disease.

The question of what would be the effect of isolated CFI overactivity could be answered by the generation of a model organism which overexpressed a constitutively active form of *cfi* (i.e. a mutation which abrogated the requirement for a cofactor) or else the identification of humans in which a gain-of-function mutation of CFI caused disease. Individuals with heterozygous mutations in CFI are well-represented in individuals and families with aHUS and C3GN (Fremeaux-Bacchi, Dragon-Durey et al. 2004; Caprioli, Noris et al. 2006; Servais, Fremeaux-Bacchi et al. 2007) but, where reported, the functional effect of these mutations has been to *inhibit* CFI protease activity.

Clearly, a defect which resulted both in alternative pathway dysregulation *and* iC3b generation would provide an attractive mechanism to explain the pathogenesis of CFHR5 nephropathy, however no data exist which support this, and the absence of C3 depletion in affected individuals argues against it.

In addition, since normal CFHR5 is not sufficient to regulate C3 turnover in the circulation in the absence of CFH and there is no abnormality of circulating complement C3 levels in individuals with CFHR5 nephropathy, it is intuitively unlikely that the primary mechanism is one of dysregulation of complement in the circulation, although clearly increased production of a C3 metabolite which has particular affinity for the glomerulus is not impossible.

4.3.2 CFHR5$^{12123\text{-}9}$: A reduced function mutation?

The data presented here demonstrate that both recombinant and patient-derived mutant CFHR5 have impaired ability to bind to heparin and complement-coated eukaryotic surfaces. However it is important to emphasize that, in these assays, CFHR5$^{12123\text{-}9}$ displayed reduced but not absent binding and should therefore be regarded as a hypomorphic rather than a complete loss-of-function allele – at least with respect to surface/heparin binding.

If the observed reduction in the binding affinity of mutant CFHR5 for host surfaces is indeed the *sine qua non* of CFHR5 nephropathy then there are three important corollaries: Firstly that CFHR5 must play a non-redundant role in the protection of glomerular surfaces, since its failure to localise there results in disease (see ***Figure 4.3-1***); secondly that administration of (wild-type) CFHR5 to people with CFHR5 nephropathy would effectively treat the disease by supplementing the deficient protein; and thirdly that in humans two normally functioning copies of the CFHR5 gene are required to prevent glomerulonephritis.

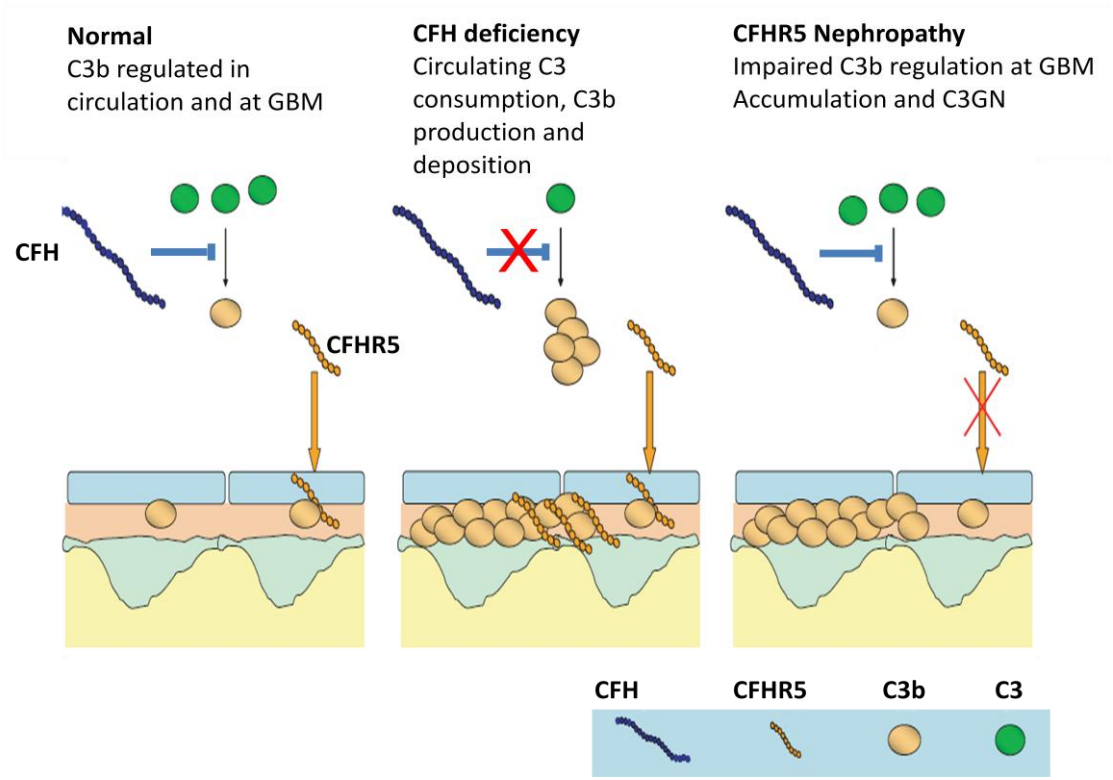


Figure 4.3-1 Proposed model for CFHR5

Showing a non-redundant role for CFHR5 in regulating C3 activity at glomerular endothelial surfaces (left panel). When CFH is deficient in the circulation (middle panel), uncontrolled C3 activation leads to runaway C3 deposition, overwhelming the ability of CFHR5 to regulate (or clear) C3 from the glomerular surfaces. When CFHR5 is deficient or mutated (right panel), circulating C3 regulation is normal but regulation at glomerular surfaces is impaired leading to disease. GBM, glomerular basement membrane.

While normal CFH function is necessary and (in the absence of a significant abnormality of C3, CFI or CFB) sufficient to regulate complement activation in the circulation (illustrated by the low C3 seen in humans and mice with CFH deficiency or depletion), both CFH *and* CFHR5 appear to be required for the normal regulation of complement activation at the glomerular surface.

If substitution of one normal allele of CFHR5 with one hypomorphic allele is sufficient to cause disease, then the normal circulating levels of CFHR5 are not in massive excess with respect to preventing C3 deposition in the glomerulus in normal humans. This is compatible with the observations that a) normal copies of CFHR5 are not sufficient to protect the glomerulus from damage in the presence of genetic or acquired causes of pathological

alternative pathway activation (for example C3 nephritic factor or complete CFH deficiency which cause DDD) and b) that activation of the classical pathway by antibodies (produced in response to infections or in autoimmune diseases such as systemic lupus erythematosus) can lead to complement deposition in the kidney and MCGN in individuals without abnormal CFHR5 alleles. In addition, the hypothesis that CFHR5 levels in healthy humans are only a little higher than that required to prevent spontaneous C3GN is not an *a priori* unlikely proposition since complement activation must remain finely balanced between the need to destroy invading microorganisms and the need to protect host surfaces (see Chapter 3); it is easy to imagine how an allele conferring *reduced* complement activation (for example a gain-of-function mutation or increased gene dose of a complement regulator) could result in death from infection during infancy, especially in the pre-antibiotic era. Following this line of reasoning, the issue of whether $CFHR5^{12123\text{-}9}$ might (in some environmental context) actually confer a survival advantage (notwithstanding the risk of kidney disease in adulthood) is discussed in Chapter 5.

The pattern of disease seen in patients with CFHR5 nephropathy is compatible with a partial impairment of the ability to tolerate activation of the immune system. The disease is characterised by episodes of macroscopic haematuria accompanied by acute deterioration in renal function which coincide with infections (often of the upper respiratory tract). It may be that, while in health complement regulatory activity of CFHR5 at the glomerular surface is sufficient to prevent disease, when complement turnover is increased owing to increased antibody production in response to infection, the complement-regulating capacity of the locally subnormal level of CFHR5 is overwhelmed and glomerulonephritis results. Consistent with this, it is recognised that mutations in or deficiency of other complement regulatory proteins (particularly CFH) most commonly produce clinical disease only in the context of an environmental immunological trigger such as a viral infection (Goodship and Kavanagh 2010).

This model has two exciting implications. The first is that, administration of CFHR5 to people with CFHR5 nephropathy would potentially correct the molecular defect, curing the disease, and might only be required during episodes of immunological stimulation (i.e. intercurrent infections). Secondly, if CFHR5 levels in genetically normal human beings are not in sufficient excess to protect the kidney in situations in which there is extreme complement activation via the classical pathway (for example lupus nephritis, IgA nephropathy or MCGN resulting from chronic infections) then CFHR5 supplementation in people with these (sporadic) diseases may provide a therapeutic benefit by enhancing the negative regulation of complement at the glomerular surface.

4.3.3 CFHR5$^{12123\text{-}9}$: A dominant negative mutation?

Since the physiological function of CFHR5 is not known, it is quite possible that CFHR5 is actually an *antagonist* of CFH and that the CFHR5$^{12123\text{-}9}$ mutation enhances this effect, either by being cleared less effectively by the kidney (owing to its more negative charge preventing transit across the glomerular basement membrane) or by more effective competition for binding to CFI. In support of this hypothesis is the observation that the cofactor activity of CFHR5 is several fold weaker than that of CFH (with CFH being effective in the 1-2 ng/µL range, compared with 300-400 ng/µL CFHR5 required for detectable cofactor activity (McRae, Duthy et al. 2005)) – it may therefore be that its physiological function is to antagonise (by competing with) CFH-mediated cofactor activity and that impaired clearance of CFHR5$^{12123\text{-}9}$ results in increased levels of an inhibitor of the negative regulation of C3 by CFH. Although conceivable, this hypothesis is rather unlikely for 2 reasons: Firstly, the site of C3 deposition in CFHR5 nephropathy is the kidney, precisely the location which CFHR5$^{12123\text{-}9}$ is less able to bind (see ***Figure 4.2-16***) rendering a dominant negative action at this site unlikely; and secondly, CFHR5$^{12123\text{-}9}$ actually has enhanced cofactor activity compared with wild type CFHR5, implying that it would it be *less* effective as a competitive antagonist of CFH.

Clearly, CFHR5 and CFHR5$^{12123\text{-}9}$ may have additional activities (perhaps relating to interactions with C5 or other complement proteins) which have not been interrogated, and it is possible that CFHR5$^{12123\text{-}9}$ has activity or activities which are completely absent in wild type CFHR5. If this is the case then inhibition of CFH by CFHR5$^{12123\text{-}9}$ would not imply that wild type CFHR5 is also capable of acting as an inhibitor of CFH.

4.3.4 CFHR5 nephropathy: lessons from therapy

The observation of recurrent C3GN post transplantation proved that the disease results from an abnormality of a circulating factor, since proteins which regulate complement which are synthesised by the allograft can be predicted to be normal.

In the two cases for which clinical data are available, use of immunosuppressive therapy which targeted the cellular/adaptive immune system (i.e. mycophenolate mofetil and cyclophosphamide) provided no evidence of benefit and in both instances was accompanied by acceleration in the rate of decline of renal dysfunction. The mechanism by which immunosuppression could exacerbate renal damage in CFHR5 nephropathy is not immediately obvious, but one possibility is that impairment of the adaptive immune system results in impaired clearance of colonizing microorganisms from the nasopharynx, skin and urinary tract. These organisms therefore are more likely to cause infections which are cleared by the innate immune system – especially complement, perhaps by the alternative and mannose binding lectin pathways. The increased complement turnover produced in this scenario would provide the substrate for increased C3 deposition in the kidney in CFHR5 nephropathy. Consistent with this hypothesis are the observations firstly that macroscopic haematuria and renal dysfunction is commonly triggered by respiratory tract infection, and secondly that in 2 individuals (both of whom were subsequently shown to have CFHR5 nephropathy) persistent macroscopic haematuria accompanying tonsillitis was terminated by tonsillectomy (Dr A Pierides, personal communication).

Although plasma exchange did seem to provide clinical evidence of benefit, with abrupt cessation of macroscopic haematuria, it should be appreciated that the creatinine was falling at the time the treatment was instituted and, although the rate of improvement in renal function appeared to be increased, this is a single therapy in a single individual. Nonetheless it is unlikely that plasma exchange (in contrast with conventional immunosuppression) had a detrimental effect. Clearly, further prospective studies are needed to determine whether this therapy offers a way of modifying the natural history of CFHR5 nephropathy.

Inhibition of C5 with the humanised monoclonal antibody Eculizumab has been shown to be effective in treating haemolysis in paroxysmal nocturnal haemoglobinuria (Parker 2009), which usually results from a bone marrow somatic mutation in the PIGA gene causing impaired expression of Decay Accelerating Factor (DAF, CD55) and other complement regulators on the surface of red cells. However, it has recently been shown that treatment with Eculizumab does not completely correct the half-life of red blood cells and does not prevent the deposition of C3 on their surfaces, suggesting mechanisms of disease in PNH which do not require C5 (Risitano, Notaro et al. 2009). Although C5-9 is detectable in the kidney in CFHR5 nephropathy, the question of whether renal injury in this disease is a consequence of C5 activation, or results from other consequences of C3 deposition has yet to be answered. In this context, while Eculizumab therapy is an attractive option (similar to its potential value in aHUS), clinical studies will be required to determine whether therapy with this agent improves outcome in CFHR5 nephropathy.

4.4 Conclusion

While the physiological function of CFHR5 is unknown, the mechanism by which heterozygosity for the CFHR5$^{12123\text{-}9}$ allele causes disease will be open to question. I have discussed 3 potential mechanisms and have taken the view that it is most likely (judging by the currently available evidence) that renal disease in CFHR5 nephropathy results from impaired targeting of the CFHR5$^{12123\text{-}9}$ protein to the glomerulus resulting in local insufficiency for a negative regulator of complement. Immunosuppression does not appear to result in significant benefit in the disease and treatment strategies which result in either normalisation of CFHR5 or inhibition of complement activation seem to hold more promise for successful treatment of CFHR5 nephropathy.

This work has raised a number of questions, the most interesting of which are as follows:

1. Biochemistry and physiological properties of CFHR5$^{12123\text{-}9}$
    a. Does CFHR5$^{12123\text{-}9}$ interact with C3b, immobilised or in the fluid phase, and how does its affinity compare with that of wild type CFHR5?
    b. What is the structure of CFHR5, alone or in complex with C3b or CFI?
    c. What is the structure of CFHR5$^{12123\text{-}9}$, alone or in complex with C3b or CFI?
    d. Is CFHR5 excreted in the urine and if so is the excretion of CFHR5$^{12123\text{-}9}$ impaired?
2. Variation in humans
    a. Do people with other mutations in CFHR5 exist, and if so what effects do these mutations have?
    b. Are genetic variants of CFHR5 differentially represented in cohorts of people with complement deposition in the kidney (e.g. those with lupus nephritis, IgAN, or MCGN) compared with controls?

c. Do people homozygous or heterozygous for complete deficiency for CFHR5 exist and if so, what is their phenotype?

d. Do levels of CFHR5 correlate with renal injury in people with autoimmune or infectious diseases associated with increased antibody production? Is CFHR5 depletion a biomarker for glomerular injury?

3. Model organisms

a. Can administration (or transgenic expression) of CFHR5 in mice deficient for cfh ameliorate renal injury and/or C3 consumption in the circulation? If so, at what dose and is this ability shared by $CFHR5^{12123\text{-}9}$?

b. Can administration of $CFHR5^{12123\text{-}9}$ (exogenous or transgenic) to mice cause renal injury?

4. Treatment of CFHR5 nephropathy

a. Does plasma exchange in people with CFHR5 nephropathy prevent injury?

b. Is inhibition of C5 protective in CFHR5 nephropathy?

c. Would depletion of C3 be protective in CFHR5 nephropathy?

d. Can administration of CFHR5 in humans with CFHR5 nephropathy, C3GN or MCGN ameliorate renal injury?

Clearly, addressing these questions will require a significant amount of further work.

Chapter 5: CFHR5 nephropathy: clinical spectrum, prevalence and distribution

5.1 Introduction

The previous chapters describe the autosomal dominant inheritance of microscopic and synpharyngitic macroscopic haematuria, C3 glomerulonephritis (C3GN) and (in affected males) renal failure occurring in 2 ostensibly unrelated families with ancestry in the Troodos Mountains of Cyprus. A genome-wide linkage study followed by haplotype analysis and molecular dissection of the linked region allowed identification of a novel, expressed mutation of the *Complement Factor H Related protein 5 (CFHR5)* gene which cosegregated with the disease. Additional work demonstrated that the mutant protein (recombinant or from patient serum) exhibited impaired ability to bind to heparin, complement-coated erythrocytes and complement-coated glomerular basement membranes. The ability of the mutant protein to degrade C3b in solution was increased compared with wild type CHFR5 and it remains unclear whether it is impaired targeting of the mutant protein to the glomerulus or increased iC3b production (or both) which leads to the disease. The possibility also existed that an undetected mutation, present on the cosegregating haplotype, was responsible for causing the disease and the CFHR5$^{12123\text{-}9}$ mutation just happens to reside on the linked chromosomal segment and is unrelated to disease pathogenesis. Notwithstanding this possibility, I considered that the CFHR5$^{12123\text{-}9}$ mutation was likely to be causative for 2 principal reasons:

1. CFHR5 was among the 6 genes identified as a candidate gene *prior* to the linkage study (in which only 0.7% of the genome was linked with the disease, and only 0.3% of the genome was inherited by affected individuals in both families from a common ancestor). The prior probability of the causative mutation residing in *CFHR5* was therefore high. Bayesian logic dictates that invoking an additional (hypothetical) mutation in some other gene (by definition not a prior candidate since all the prior candidates were sequenced before the linkage study) is very much less likely to explain the data than is the detected CFHR5 mutation.

2. The mutation was associated with the expression of a novel CFHR5 variant which had altered function (i.e. reduced binding to the glomerular surface) in a way compatible with causing the disease.

Clinical genetics is essentially a probabilistic science, inasmuch as proof (to a level comparable with fulfilment of Koch's postulates in microbiology) is seldom possible. Specifically, to prove that a mutation causes a human disease beyond *all* doubt would require specific, targeted introduction of the mutation in a human being of known genetic background (for example one of identical twins). This is an undertaking which is beyond the capability of the current state of the art and is, in any case, unlikely to be ethically acceptable under any foreseeable circumstances. Instead, there are accepted to be several types of evidence which can lead to a novel genetic variant which is linked with a disease being accepted (i.e. proven beyond *reasonable* doubt) to be responsible for causing a disease.

1. The identification of a similar genetic variant in the same gene causing the disease in an unrelated individual: while not completely proving the case, this would raise the likelihood that the disease is caused by a mutation in the gene in question by several orders of magnitude. In the context of CFHR5 nephropathy, ongoing work aims to sequence CFHR5 in other individuals with C3GN, but since this is such a rare disease (fewer than 40 cases previously described in the world literature) this approach appears to have a rather low probability of resolving the issue in a realistic timeframe.
2. The identification of individuals inheriting a mutation from the same ancestor with recombinations flanking the mutation closely enough to exclude inheritance of all other genetic variants – in other words if the common haplotype is small enough to be sequenced in its entirety then mutations in other genes in the interval can be excluded, proving the case by elimination. Conventionally, this has been viewed as likely only to occur in mutations which have been present in a population for a great many years (i.e.

usually autosomal recessive conditions), however advances in sequencing technology have meant that sequencing larger genomic regions is becoming increasingly practical as a solution to this problem.

3. A clear demonstration, using recombinant proteins, that the mutation affects the function of the protein sufficiently to provide a molecular mechanism for the disease. In the case of whole-gene deletions or early premature termination codons, or where the function of the gene is well understood, this is readily achievable (see Chapter 2: HIF2α erythrocytosis and pulmonary hypertension), but where the function and biochemical interactions of the wild type protein are less well understood and the altered structure and function of the gene resulting from the mutation are not easily predictable, this may be difficult to achieve. The physiological role of CFHR5 is not understood, and the homology with other members of the CFH/CFHR1-5 gene cluster adds significant complexity to this undertaking. It is likely that for this approach to resolve the issue for CFHR5 nephropathy, a range of unbiased assessments of the biochemical and physiological functions of the protein *in vitro* and *in vivo* will be necessary.
4. Introduction of the mutation in the orthologous gene in an animal model causing a phenotype sufficiently similar to the human disease. The level of proof that this type of experiment provides is crucially dependent on the similarity of the model organism to humans in relation to the system being interrogated.

Given that the approaches likely to prove beyond reasonable doubt that the mutation identified in CFHR5 was responsible for the disease were beyond the time and resources available for this project, I therefore considered the evidence that would *disprove* this hypothesis. Although the $CFHR5^{12123\text{-}9}$ allele was not detected in control individuals from the UK population, the possibility remained that it was a common variant in Cyprus: demonstrating its presence in large numbers of unaffected individuals in Cyprus would cast significant doubt on

its role as the pathogenic mutation in CFHR5 nephropathy and prompt the search, within the segregating haplotype, for another putative causative genetic change. Conversely, it seemed possible that identification of further (unrelated) individuals with the mutation who also had the disease would support the hypothesis and reduction of the size of the shared haplotype may allow proof by elimination (see 2 above) of the pathogenic role for the mutation.

The observation that the disease was cosegregating in two ostensibly unrelated families raised the possibility that it may be present in other individuals in Cyprus. Since the onset of clinically significant renal impairment was after the 4th decade in all individuals, and because the mutation is predicted to increase the activation of the complement system, it did not seem impossible that the mutation (perhaps in a certain environmental context) could actually enhance survival by increasing the ability to resist infection by pathogenic micro-organisms in childhood.

In this chapter I will present data demonstrating a high prevalence and wide distribution of the mutation within Cyprus and show that it has probably been present in the Cypriot population for at least two millennia.

5.2 CFHR5$^{12123\text{-}9}$ in the Cypriot population

5.2.1 Estimation using healthy controls

In order to measure the frequency of the allele in the general Cypriot population accurately (i.e. without introducing bias for inclusion of people with renal disease) I aimed to identify a collection of appropriate DNA samples from the Cypriot population. In order to achieve this, I performed a pubmed search using the terms "Cyprus" and "DNA." This identified a recent paper (Loizidou, Michael et al. 2008) describing the results of a breast cancer study (the "MASTOS" study) which used 1109 Cypriot women and 1177 age-matched healthy controls. I contacted Dr Andreas Hadjisavvas, the corresponding author, and he agreed to screen 1015 of the control subjects for the presence of the allele using PCR primers which I sent to him. This experiment identified a single individual (out of 1015) who carried the allele – a finding which I confirmed by (blinded) analysis of 15 samples (including the individual in whom he had detected the mutation) from this cohort which he provided (***Figure 5.2-1***).

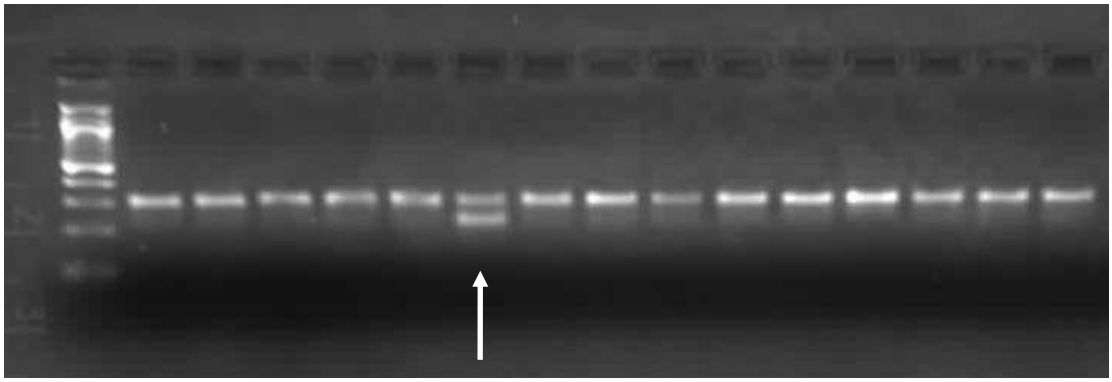


Figure 5.2-1 CFHR5 duplication in a control individual
3-primer PCR reaction diagnostic for CFHR5 duplication (see Chapter 3: Identification of CFHR5 nephropathy) in 15 Cypriot individuals from the MASTOS study. A single individual (arrow) was found to be affected among 1015 Cypriot control individuals (1000 unaffected individuals not shown).

Under the terms of the ethical approval under which the MASTOS study was performed, it was not possible to obtain any further clinical or demographic details for this single anonymous

affected individual. Nevertheless, I concluded firstly that the duplication is not a common variant in Cyprus and secondly that at least one additional affected family was likely to exist.

Although identification of only a single individual in a sample of ~1000 does not allow a robust estimation of the actual frequency in the population, EB Wilson (Wilson 1927) proposed a method of estimating a 95% confidence interval of the true frequency from an observed proportion (Newcombe 1998). Using an online statistical calculator (VassarStats 2010) to calculate this yielded an interval of 0.0002 to 0.0056 (equating to 1:5000 to 1:179) for the estimated proportion of the Cypriot population who harbour the mutation. This suggested that, rather than being a vanishingly rare condition affecting only those individuals already identified, CFHR5 nephropathy might be a relatively common genetic condition in Cyprus. For comparison, estimates of the frequency of adult polycystic kidney disease in Caucasian populations range from 1:1000 (Dalgaard 1957) to 1:2500 (Davies, Coles et al. 1991).

Clearly, refining this estimate using this method to improve the statistical certainty of the allele frequency would involve sampling a very large number of the general Cypriot population, an undertaking which would require significant resources. I therefore pursued alternative approaches to estimate the frequency of the allele and the disease in Cyprus.

5.2.2 Estimation using family data

In order to predict how likely it was that additional mutation carriers would exist, I attempted to estimate the allele frequency in the local population from which Family 1 and Family 2 were drawn. In order to do this I first estimated how many generations were likely to have separated the families. Since the families could not trace their ancestry to a single ancestor genealogically, this implied that at least 4 generations separated them (unless more recent undocumented extra-pair paternity had occurred) and further information was unlikely to be gained from documentary evidence. However, the size of the shared haplotype between

the families could be used to estimate how many generations had elapsed, following the reasoning of Genin et al (Genin, Tullio-Pelet et al. 2004):

5.2.2.1 Minimum age of the allele

The genetic distance between 2 loci is determined from the recombination fraction θ (which is the observed proportion of meioses resulting in a recombination between the two loci) according to a mapping function, such as that developed by Kosambi (Kosambi 1944):

$$S = \frac{1}{4}\ln\left(\frac{1+2\theta}{1-2\theta}\right)$$

where *S* is the distance in hundreds of cM (abbreviated to M below).

For a distance *S* M, the expected recombination frequency, θ_S, is given by the inverse of the mapping function:

$$\theta_S = \frac{\frac{1}{2}(e^{4S}-1)}{e^{4S}+1}$$

It follows that the likelihood that no recombinations occurred between a mutation and a marker *x* which is *k* M distant in *n* meioses is given by:

$$T(x) = (1-\theta_k)^n$$

The likelihood that a recombination occurred between marker *x* and the adjacent marker *x-1* is given by:

$$F(x) = T(x-1) - T(x)$$

If marker *x* is the marker in the haplotype shared between members of both families most distant (e.g. farthest centromeric) from the mutation, it follows that either one or both individuals (from the two different families) must have inherited a recombination in this region

in the *n* generations since their most recent common ancestor. Therefore, the likelihood that this occurred is given by:

$$K(x) = F(x)^2 + 2T(x)F(x)$$

and the likelihood of the observed haplotype being transmitted from an ancestor *n* generations ago is given by combining the likelihood for the centromeric- AND telomeric-most shared markers:

$$L(n) = K(x_{centromeric}) \times K(x_{telomeric})$$

The value of *n* for which *L(n)* is maximized is therefore an estimate of the number of generations which has elapsed since the most recent common ancestor of individuals sharing a common haplotype. Clearly, *L(n)* tends towards 0 as *n* becomes large.

Plotting the number of generations against the likelihood of the observed haplotype allowed estimation of the most likely number of generations to have passed since the most recent common affected ancestor of the individuals from the two families ($MRCA_{families}$) (***Figure 5.2-2***). This yielded a figure of 10 generations, equating to roughly 250 years, which was therefore the likely minimum age of the mutation (clearly the mutation could have arisen earlier – prior to the divergence of families 1 and 2).

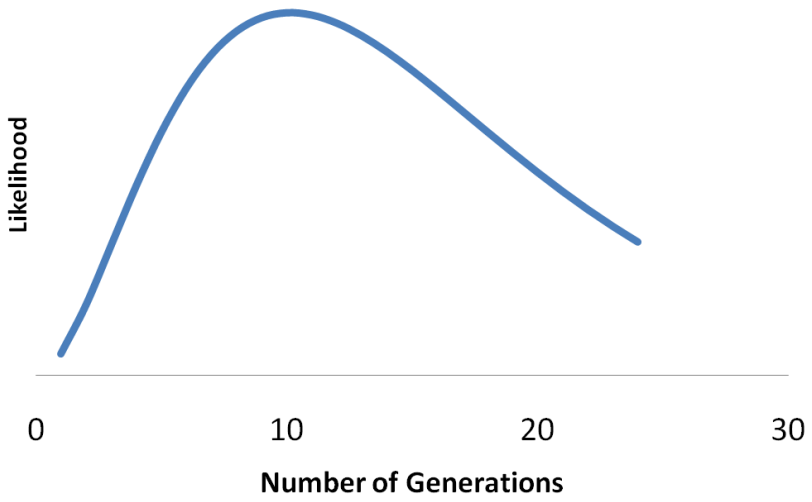


Figure 5.2-2 Estimated time since most recent common ancestor of families 1 and 2

Genin et al also determined 95% confidence interval for this estimation using Bayes' theorem, where the posterior probability of the observed haplotype being shared across *m* generations is L(m) divided by the sum of all the likelihoods for all other number of generations:

$$P(m) = \frac{L(m)}{\sum_{u=1}^{r} L(u)}$$

Where *r* is very large (and hence *L(r)* is very small). The 95% confidence interval is determined by finding the values for *nlow* and *nhigh* for which $\sum_{u=1}^{nlow} P(u) = 0.025$, and $\sum_{u=1}^{nhigh} P(u) = 0.975$. This yields a 95% confidence interval of between 3 and 32 generations.

5.2.2.2 Likelihood of further affected relatives existing

The valley containing Gerakies and Kalopanagiotis also contains 3 other villages: Moutoullas (which is contiguous with Kalopanagiotis), Oikos and Pedoullas (***Figure 5.2-3***). The combined populations of these villages, according to the 2001 census, is 1081 (Cyprus Ministry of Finance 2001). The past thirty years has seen a reduction in the population of the valley – the

population of the valley fell by approximately 25% between 1992 and 2001 (Cyprus Ministry of Agriculture Natural Resources and Environment 1991) and it is likely that the pre-1950 population was of the order of 2000 people.

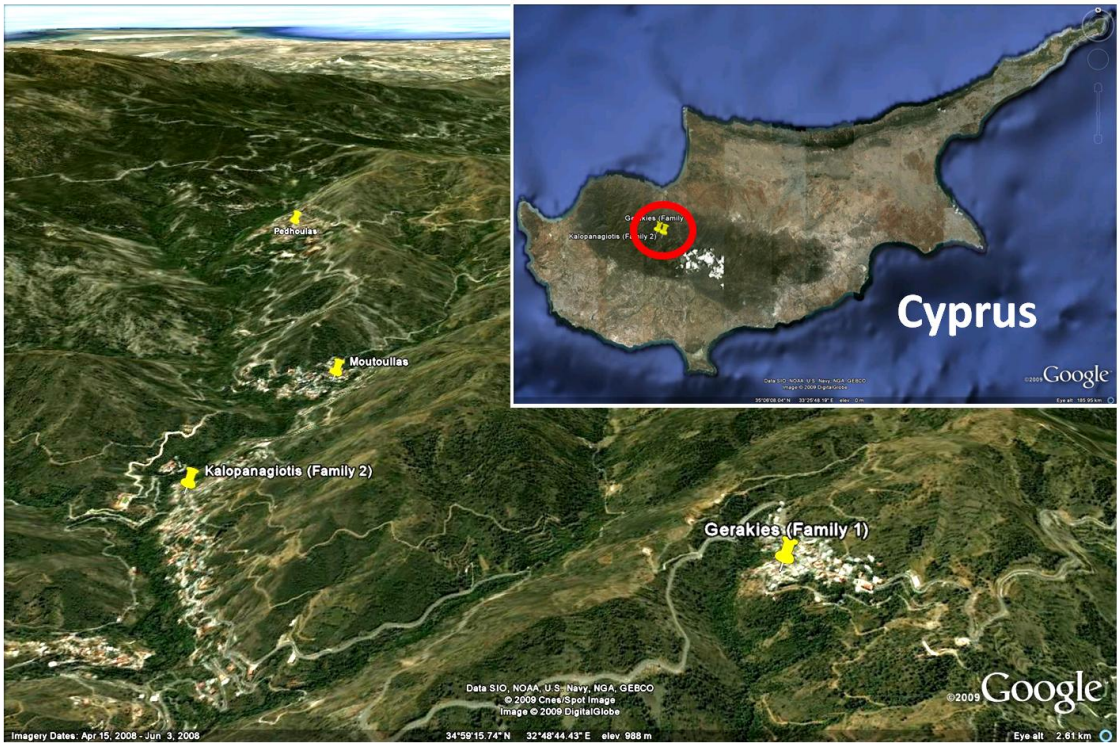


Figure 5.2-3 Satellite image of the valley in the Troodos mountains
Families 1 and 2 reported ancestry from this valley. The position of the valley on the island of Cyprus is shown (inset). Images downloaded from www.google.com.

In an isolated, randomly mating population, as genealogies are traced back in time there comes a point when every currently existing member of the population is descended from a single individual – this individual is termed the Most Recent Common Ancestor ($MRCA_{population}$). There may be more than one MRCA alive at this point, and with each generation one goes back beyond this the number of common ancestors (as a proportion of the population) increases. This is in part because *all* the ancestors of the MRCA must be common ancestors of the whole present-day population. Eventually, a generation is reached in which *all* the individuals in this generation are either ancestors of everybody living today or else they left no descendants – this is the Identical Ancestors (IA) point.

It has been shown that, for a randomly mating population of size *n,* the time since the $MRCA_{population}$ is sharply distributed about $\log_2(n)$ generations in the past, with the number of generations since the IA point distributed about 1.77 x $\log_2(n)$, at which time the probability that a randomly chosen individual is a common ancestor of the *entire* present population is approximately 0.8 (Chang 1999).[a]

In order to make some estimate of whether it is likely that other affected families exist, I made several assumptions:

1. The effective population, *n*, of the valley was approximately 2000 people
2. That random mating occurred within the valley – anecdotal reports from the family support this assumption in general, however it should be noted that the model proposed by Chang takes no account of the fact that an individual usually produces offspring with a single partner multiple times, rather than each offspring being the result of an independent random mating event. Nonetheless, for small family sizes the approximation is likely to be valid
3. Reproductive fitness of individuals harbouring the mutation is not significantly reduced compared with the general population. I viewed this as a reasonable assumption because symptomatic or morbid renal impairment had not been observed in any individuals younger than 40 years of age
4. Migrants into the valley did not contribute much to the gene pool. This is supported by the observation that, historically, migration has been *out* of the valley. The notable exception to this is the presence of the monastery (discussed in Chapter 3) and this issue clearly places a limit on the validity of this estimation since the model takes no account of an influx of new founders

[a] A startling corollary of this is that, given that he is known to have living descendants, Charlemagne (742-814) is predicted to be an ancestor of *all* people of recent Western European descent

5. The MRCA of families 1 and 2 ($MRCA_{families}$) resided in the valley approximately 10 generations ago

It can therefore be estimated that, while the MRCA of the whole valley ($MRCA_{valley}$) is likely to have lived approximately 11 generations ago, and that 80% of the population 19 generations ago were ancestors to the whole of the current population, the $MRCA_{families}$ lived more recently than this and hence is unlikely to have been ancestral to the whole valley. However, considering that the $MRCA_{families}$ likely lived so long ago (i.e. relatively close in time to the IA point) and had given rise to at least 2 families in existence now, there appeared to be at least a reasonable chance that other descendants of this individual would exist. This result is intuitively correct since even people from societies with small family sizes can be expected to have a very large number of 9^{th} cousins. Furthermore, there is no evidence to suggest that the $MRCA_{families}$ was the first individual to harbour the mutation and clearly with each preceding generation the likelihood increases that an affected individual was ancestral to a greater proportion of the current residents of the valley.

Since humans are diploid organisms, the probability of inheriting a particular allele from an ancestor *within a randomly mating population* depends on the frequency of that allele in the population (because descent within a finite population can be by more than one possible intermediate in each intervening generation – contrasting with the completely outbred situation where the probability of inheriting a given allele from a given ancestor is $1/2^n$ where n is the number intervening of generations), so while having an affected ancestor is *necessary* it is certainly not *sufficient* to result in an individual being affected. In conclusion, it appeared likely that further individuals from the valley with the *CFHR5* duplication would be identified if an appropriate population were tested. The obvious questions were firstly whether such individuals manifested renal disease, and secondly what the frequency of the allele was.

5.2.3 Identification of additional affected individuals

This reasoning supported the hypothesis that further investigation of the Cypriot population was likely to identify other people with the disease. I therefore contacted Dr Alkis Pierides, a nephrologist in Cyprus who had looked after the index case from family 2 (see Chapter 3) and who, in collaboration with Professor Constantinos Deltas, had studied inherited renal disease, identifying mutations in the *Col4A3* and *Col4A4* genes segregating with autosomal dominant inheritance of microscopic haematuria and focal segmental glomerular sclerosis (Voskarides, Damianou et al. 2007).

Dr Pierides and colleagues used the diagnostic PCR reaction to screen a cohort of 84 Cypriot patients with advanced or end-stage chronic renal disease. 23 of these individuals had presumed glomerulonephritis with no histological diagnosis; 19 had biopsies reported as membranoproliferative or mesangioproliferative glomerulonephritis or focal sclerosis; and the remaining 42 patients had end-stage renal failure of unknown cause. This identified 1 female and 3 male patients who harboured the duplication, proving the supposition that other individuals with the allele would be identified, and additionally extending the evidence for an association of the allele with renal disease. None of these individuals initially reported descent from the Troodos region.[a]

In addition, 2 families were identified in which the duplication cosegregated with microscopic haematuria (***Figure 5.2-4***). One of these families, designated family 3, was identified from a proband in the cohort described above. In addition to microscopic haematuria, individuals from these families exhibited synpharyngitic macroscopic haematuria and/or renal impairment. Neither of these additional families could trace their ancestry to the Troodos Mountains and one of them (family 4) reported ancestry in Deftera which is 12km southwest of Nicosia in the Mesaourian Plain.

[a] It subsequently came to light that one of these individuals could, in fact, trace their ancestry to the Troodos region.

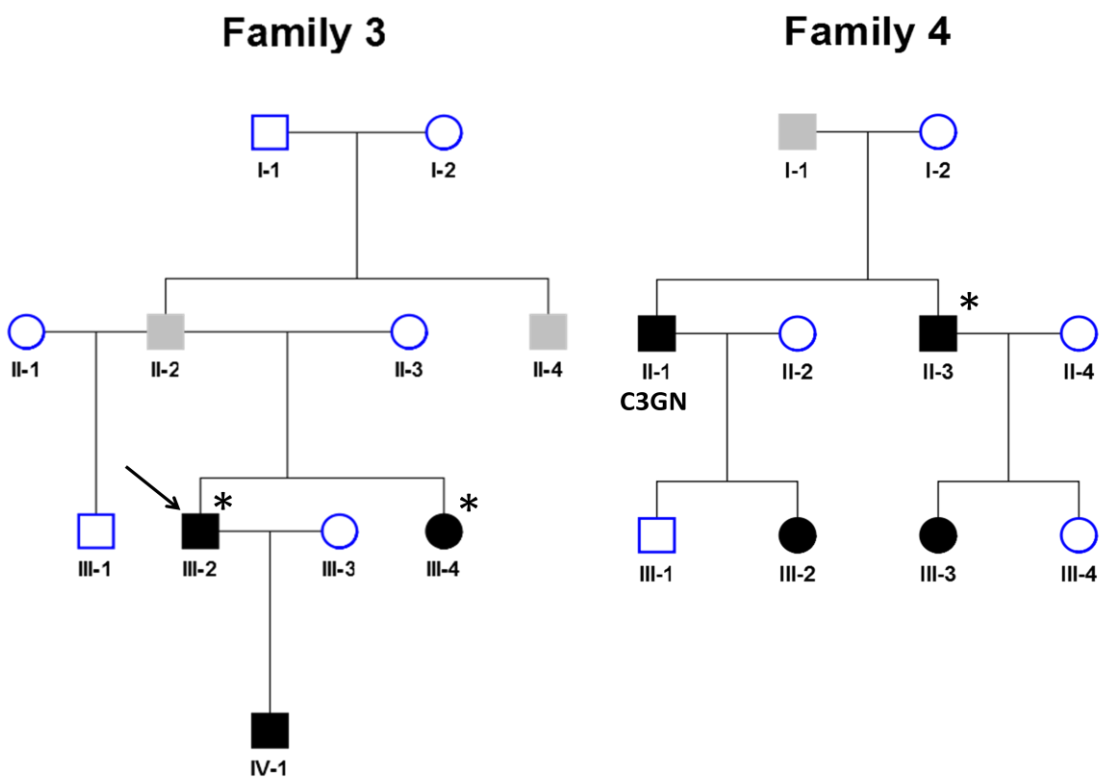


Figure 5.2-4 Two additional Cypriot families
In both families the CHFR5 duplication cosegregated with microscopic haematuria and/or renal failure (shaded black). Asterisks represent individuals who gave a history of synpharyngitic macroscopic haematuria. No DNA was available from some individuals who reported renal disease (shaded grey).

Having recognized this as a potential new endemic disease, I identified further affected individuals, all of whom were of Cypriot descent and resident in London. The first of these people was a 54 year old man undergoing haemodialysis at the West London Renal and Transplant Centre (denoted 115301). He had first noticed synpharyngitic macroscopic haematuria more than 30 years ago, with recurrent episodes approximately once every 1-2 years. He had been referred to the nephrology service at St Mary's Hospital in 1993 and underwent a renal biopsy in 1998. Although initially reported as mesangiocapillary glomerulonephritis type I, review of this biopsy demonstrated that the clinicopathological diagnosis was C3GN: there was immunostaining for C3 but not C1q or immunoglobulins. Electron microscopic examination demonstrated mesangial, elongated subendothelial and very

occasional subepithelial basement membrane dense deposits. He developed end stage renal disease in 2003 and received a cadaveric transplant in 2009 after 6 years of haemodialysis.

His brother (designated 115302) was aged 59 and had been noted to have microscopic haematuria 10 years previously, with normal appearances on cystoscopy and ultrasound imaging of the renal tract. He did not recall any episodes of macroscopic haematuria. The blood pressure and physical examination were normal and there was no detectable proteinuria. Serum creatinine was 79 µmol/l and complement C3 and other blood tests were within the normal ranges, with the exception of a serum cholesterol level of 7.5 mmol/l. The *CFHR5* duplication was identified in both these individuals (***Figure 5.2-5***) and Western blotting identified the mutant and wild type CFHR5 proteins in their circulation. While their mother (who was found not to harbour the mutation) had been born in Cyprus to Armenian parents, their father was from Moutoullas, which is a village in the Troodos Mountains which is almost contiguous with Kalopanagiotis (***Figure 5.2-3***). He had died at the age of 77 from a ruptured aortic aneurysm but had been noted to have chronic kidney disease prior to his death.

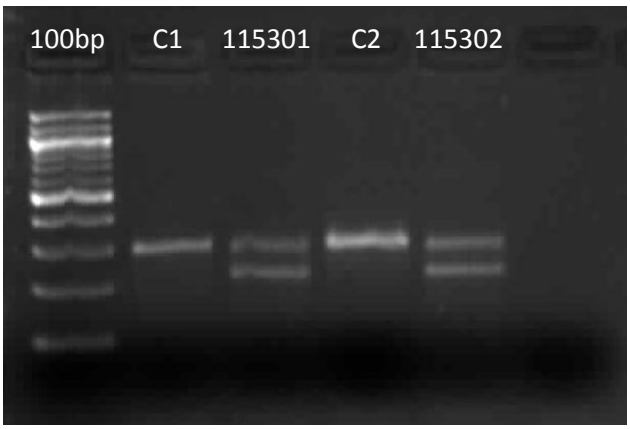


Figure 5.2-5 Diagnostic PCR reaction in 2 brothers
Demonstrating the presence of the CFHR5 mutation in 2 brothers (115301 and 115302). Both resided in London and although reported paternal ancestry from the Troodos region, were unrelated to the families 1-4. C1 and C2: unrelated controls.

Next, I identified an additional patient (designated 116301) who was undergoing regular haemodialysis at the Royal Free Hospital in London. He had noticed synpharyngitic macroscopic haematuria while a teenager and had undergone a renal biopsy at the ages of 17 and 27. These showed mesangiocapillary glomerulonephritis (with only slight mesangial hypercellularity) and positive immunostaining for C3 in the capillary walls and absent immunostaining for immunoglobulins or C1q. Electron microscopy demonstrated subendothelial electron dense material within the glomerular basement membrane and in mesangial regions. There were areas of reduplication of the basement membrane and mesangial cell interposition. The second biopsy demonstrated increased mesangial sclerosis and interstitial fibrosis when compared with the first one. He also reported abdominal pain associated with altered bowel habit. Scintigraphy indicated increased uptake of gallium-67 in the colon and colonoscopy revealed multiple superficial, discrete, white ulcers but no polyps. Biopsy specimens demonstrated increased inflammatory cells and a single focus of cryptitis. There was no evidence of granuloma, infection or dysplasia and no firm diagnosis was reached. His renal function subsequently deteriorated, with a creatinine of 140 µmol/l at the age of 32. There was an abrupt decline in his renal function at the age of 34 and he was treated with oral prednisolone and mycophenolate mofetil for several weeks. There was no apparent clinical effect on his renal impairment and steroids and immunosuppression were withdrawn. He started regular haemodialysis at the age of 35. Genetic testing demonstrated the presence of the *CFHR5* duplication and heterozygosity was confirmed on Western blotting. There was no family history of renal disease although his father had died young from intracranial malignancy. Both parents were from Nicosia and could not trace ancestry to the Troodos Mountains. Clinical and laboratory data from the first 24 people with the mutation identified are shown in ***Table 5.2-1***.

	Total (data available)			Percent		
	Male	Female	All	Male	Female	All
CFHR5 mutation detected	12	12	24	50	50	-
Mutation carriers found to have microscopic haematuria	10 (10)	11 (11)	21 (21)	100	100	100
Obligate carriers with CKD	3	0	3	-	-	-
Confirmed and obligate affected individuals with CKD*	14 (17)	2 (12)	16 (29)	82	17	55
Affected individuals with synpharyngitic macroscopic hematuria	6 (9)	2 (3)	8 (12)	67	67	67
Biopsy in affected individuals showing C3GN	6 (6)	3 (3)	9 (9)	100	100	100
Mesangial matrix expansion	6 (6)	2 (3)	8 (9)	100	67	89
Increased glomerular cellularity	6 (6)	2 (3)	8 (9)	100	67	89
Segmental capillary wall thickening	6 (6)	2 (3)	8 (9)	100	67	89
Glomerular staining for C3	6 (6)	3 (3)	9 (9)	100	100	100
Glomerular staining for C1q, IgA, IgG or IgM	0 (6)	0 (3)	0 (9)	0	0	0
Subendothelial electron dense deposits	6 (6)	3 (3)	9 (9)	100	100	100
Mesangial electron dense deposits	6 (6)	3 (3)	9 (9)	100	100	100
Scanty subepithelial electron dense deposits	4 (6)	3 (3)	7 (9)	67	100	78
Heterozygous for $CFHR5^{12123-9}$ (serum Western blot)	8 (8)	3 (3)	11 (11)	100	100	100
Normal serum C3, C4, CFH	9 (9)	4 (4)	13 (13)	100	100	100
Negative for C3NeF	3 (3)	3 (3)	6 (6)	100	100	100
Normal retinal appearance	1 (1)	1 (1)	2 (2)	100	100	100

Table 5.2-1 Clinical and biopsy features in 24 individuals with CFHR5 nephropathy
CKD, Chronic kidney disease.
"Obligate affected" refers to individuals with confirmed affected off-spring and other relatives.
*p=0.0007, comparing males with females using Fisher's exact test

5.2.3.1 Clinical features in 103 affected individuals

Multiple further individuals with the disease were then identified by screening families and individuals with microscopic haematuria and or renal impairment and C3GN. Together with Dr Pierides and colleagues, we assembled details of 103 individuals with the mutation, all either living in Cyprus or of recent Cypriot descent living in London. Ancestry was not confined to the Troodos region in this cohort (see ***Figure 5.2-6***). Of these 103 people with the mutation, 93 had evidence of haematuria on urinalysis, in 57 of these people this was the only abnormality. 36 individuals exhibited haematuria and proteinuria of whom 32 (89%) had renal impairment (including 19 people with end stage renal disease, ESRD). Renal biopsies had been performed in 18 individuals (including 2 females with haematuria alone) from 10 different kindreds and in all cases showed features consistent with C3GN. In an individual from one large family, the biopsy also showed thinning of the glomerular basement membranes.

Figure 5.2-6 Geography of CFHR5 nephropathy

Satellite image of Cyprus showing locations of reported ancestry of the first 17 families with CFHR5 nephropathy to be identified. Although showing evidence of geographic clustering around the Troodos and Mesaourian (Nicosia) region, it is not clear to what extent this represents the effect of ascertainment.

The sexual dimorphism with respect to proteinuria and renal impairment observed in the 2 initial families (see Chapter 3) was also apparent in this cohort: in affected individuals aged between 31 and 50 years, 41% males but only 9% females exhibited proteinuria and renal impairment. This rose to 69% of men and 14% women aged 51-70 and 80% men and 20% women aged >70 years. In total, 26 out of 54 males and 6 out of 49 females had renal impairment, with 15 men and only 4 women having reached ESRD. This apparent difference in susceptibility between males and females is statistically significant ($p=0.0001$ comparing males with females using Fisher's exact test) and I concluded the sexual dimorphism represents a real biological effect, although the biology underlying this is not obvious.

5.2.3.2 Minimum prevalence of the CFHR5 nephropathy in Cypriots

The total population of Cyprus is 800,000 and the Cypriot population of London is estimated to be approximately 80,000. Based on the cases already identified, the minimum possible prevalence of the mutation among the Cypriot community is 1:8500. It seems likely that the true prevalence will be somewhat higher than this since additional affected individuals are likely to exist.

5.2.3.3 Geographical distribution of CFHR5 nephropathy

The identification of multiple ostensibly unrelated individuals and families with the disease demonstrated that the disease is endemic in Cyprus and confirmed the existence of a substantial at-risk population there and in London. Having identified the disease only in people of Cypriot descent, a key question was what the age and distribution of the mutation might be. A secondary question was whether the mutation occurred within Cyprus (possibly in the Troodos region) or whether it was imported into Cyprus from elsewhere. This was potentially important because screening for the disease in other Eastern Mediterranean populations could be undertaken if a wide geographical distribution was anticipated.

I proceeded using two approaches: the first was to define, as precisely as possible, the common haplotype shared by the different individuals and families with the disease. This would allow firstly a refined estimation of the minimum age of the mutation and secondly might provide more information about the geographical distribution of affected individuals within Cyprus. The second approach was to identify the frequency of the disease-linked haplotype in the Cypriot population. Identification of a haplotype flanking the mutation which is rare in the general Cypriot population would suggest an origin outside Cyprus with migration into the island of an affected individual or cohort. Conversely, if the mutation existed within a haplotype that is common among the Cypriot population this would imply that the mutation occurred within Cyprus, or within a population ancestral to a large proportion of Cypriots.

Clearly, the mutation could also have occurred by chance on a rare haplotype within Cyprus, but the prior probability of this would be correspondingly low.

I therefore genotyped 300,000 SNPs across the genome (using the Illumina HumanCytoSNP-12 panel[a]) in 3 individuals from each of the first 4 families to be identified; individuals 115301 and 115302 (family 5) and one individual from each of the next 12 families family for whom adequate DNA was available. Families 1 and 2 are described in detail in Chapter 3 and they, along with family 5 and 4 further families, reported ancestry in the Troodos region. Families 3 and 4 reported ancestry from the Mesaourian plain (near Nicosia) along with 6 other families and the remaining 2 families reported ancestry in Chloraka (near Paphos) and Larnaca and Famagusta (see ***Figure 5.2-6***).

Where single individuals were genotyped from families it is not possible to determine the gametic phase. However, since the purpose of this experiment was to determine the maximum possible shared haplotype, identifying homozygosity for different alleles of a given SNP between 2 individuals (termed 'conflicting homozygosity') is incompatible with inheritance of a shared haplotype at that locus by those individuals. I therefore defined the shared haplotype as the region for which no conflicting homozygosity occurred between the families in question (see Appendix A: Conflicting Homozygosity analysis).

Comparison of genotypes in this region in affected individuals from 17 families firstly defined the haplotype flanking the mutation and secondly demonstrated that the minimum possible common haplotype (i.e. the genotypes across which no conflicting homozygosity occurred) comprised 127 SNPs across 1.07 cM. The recombinations defining the boundaries of this common haplotype occurred in the ancestry of family 10 (Kalopanagiotis 2) and family 14 (Lakatamia 1). In the absence of shared ancestry at this locus, the likelihood of observing no

[a] It was more cost-effective to run a commercially available chip to genotype 300,000 SNPs across the whole genome in each person than it would have been to perform targeted genotyping for 100 or so SNPs across the region of interest.

conflicting homozygosity (in just these 2 families) across this distance would be $(1-0.059)^{127} = 4 \times 10^{-4}$. Observing no conflicting homozygosity across this region in all 17 families tested is clearly vanishingly unlikely in the absence of shared ancestry. This interval therefore defines the *maximum* possible size of the haplotype shared by all the families tested.

The haplotypes are represented graphically in ***Figure 5.2-7***. This analysis also allowed inference of common sub-ancestry of some of the families, since they shared private haplotypes (implying shared ancestral recombination events) in which there is no conflicting homozygosity. Geographic clustering of families sharing common recombination events was observed, with 'private' haplotypes segregating in 3 families from the Troodos region and multiple families from Mesaouria.

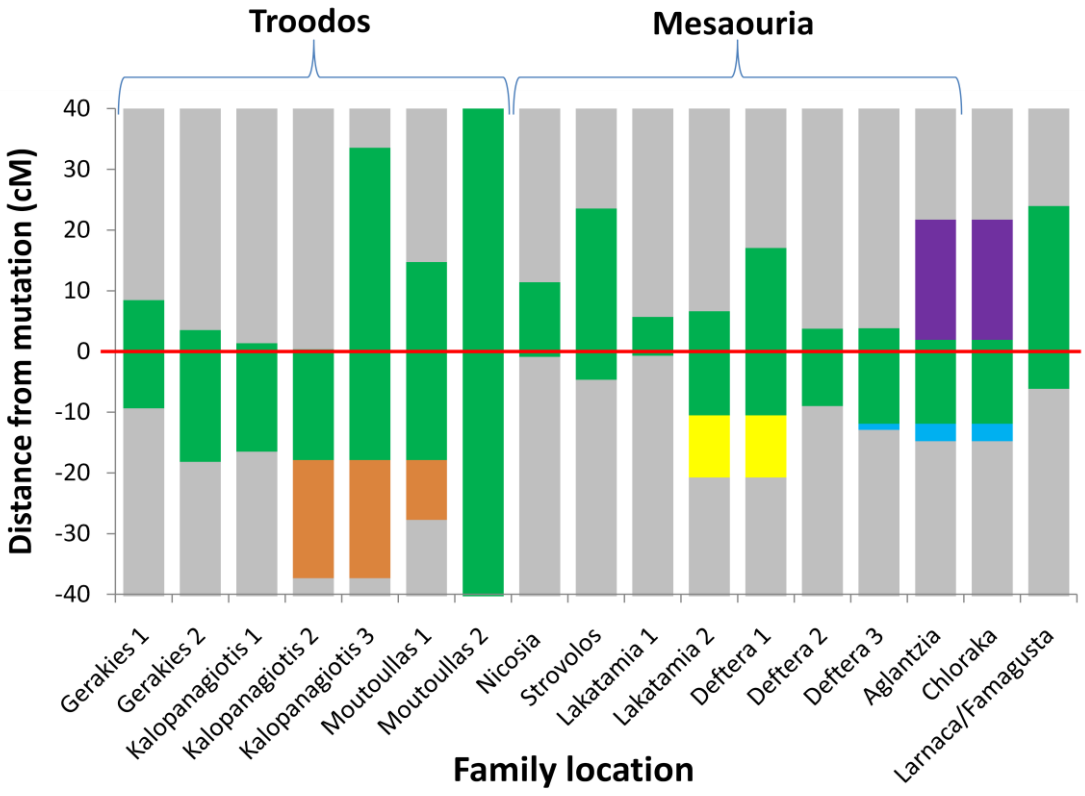


Figure 5.2-7 Common CFHR5 nephropathy haplotype
A shared haplotype (green) was present in all families harbouring CFHR5 mutation studied. In addition, private flanking shared haplotypes (orange, yellow, purple and blue bars) existed between subsets of the families, consistent with their reported geographic ancestry.

The size of the shared haplotype allowed refinement of the estimate of the number of generations since the MRCA of the most distantly related families to 93 generations (***Figure 5.2-8***, 95% confidence interval 29 to 334 generations), representing a most likely minimum age of the mutation of approximately 2500 years (95% confidence interval 725 to 8350 years).

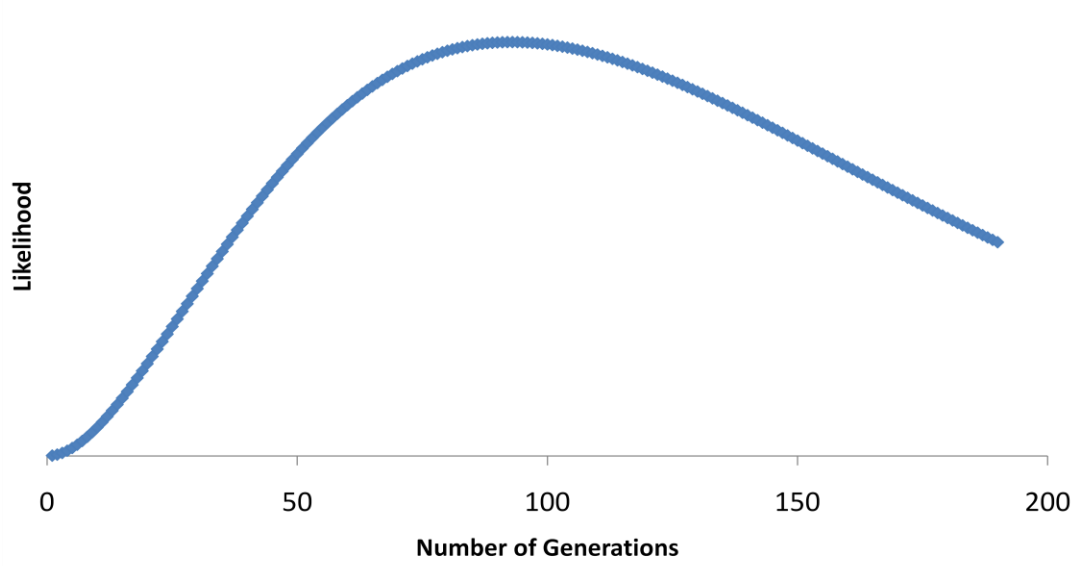


Figure 5.2-8 Estimated time since the most recent common ancestor of 17 families
The likelihood estimate was maximised at 93 generations (~2500 years)

5.2.3.4 Cypriot population structure

In order to understand better the population structure in Cyprus, and in particular to estimate whether the Troodos mountains region was genetically isolated from Mesaouria, I obtained samples of DNA which had been collected by Dr Andrie Panayiotou and Prof Andrew Nicolaides as part of a programme to investigate genetic factors contributing to cardiovascular risk. The collection consisted of 778 DNA samples: 427 from the village of Pedhoulas, in the Troodos mountains (within 5km of Moutoullas) and 247 from Nisou on the Measourian plain (***Figure 5.2-6***).

I first tested for the CFHR5 duplication in all 778 samples and did not identify it in any of these individuals (***Figure 5.2-9***). Next, haplotype-tagging SNPs spanning a 330 kbp region which included *CFH* and the *CFHR1-5* genes were genotyped in this collection, and also in affected individuals from the 4 families. Haplotype tagging SNPs were selected using the Hapmap CEU database (North American Caucasians of Northern and Western European ancestry) genotypes using HAPLOVIEW (Barrett, Fry et al. 2005) and genotyping was performed by Kbioscience Ltd (Essex, UK) using KASPar, a competitive allele-specific PCR with 3'-5' exonuclease-deleted DNA polymerase and a fluorescence resonance energy transfer quenching reporter oligonucleotide. The resulting genotypes were assembled into haplotypes using HAPLOVIEW. As would be anticipated in identical-by-descent inheritance of the mutation, the same haplotype was present in all affected families tested. This haplotype was common in Cyprus, occurring with a frequency of 0.128 (compared with 0.075 for the CEU population, see ***Figure 5.2-10***). This haplotype was not more common in individuals from Pedhoulas (Troodos) compared with Nisou (Mesaouria) (p=0.6, Cochran-Armitage trend test).

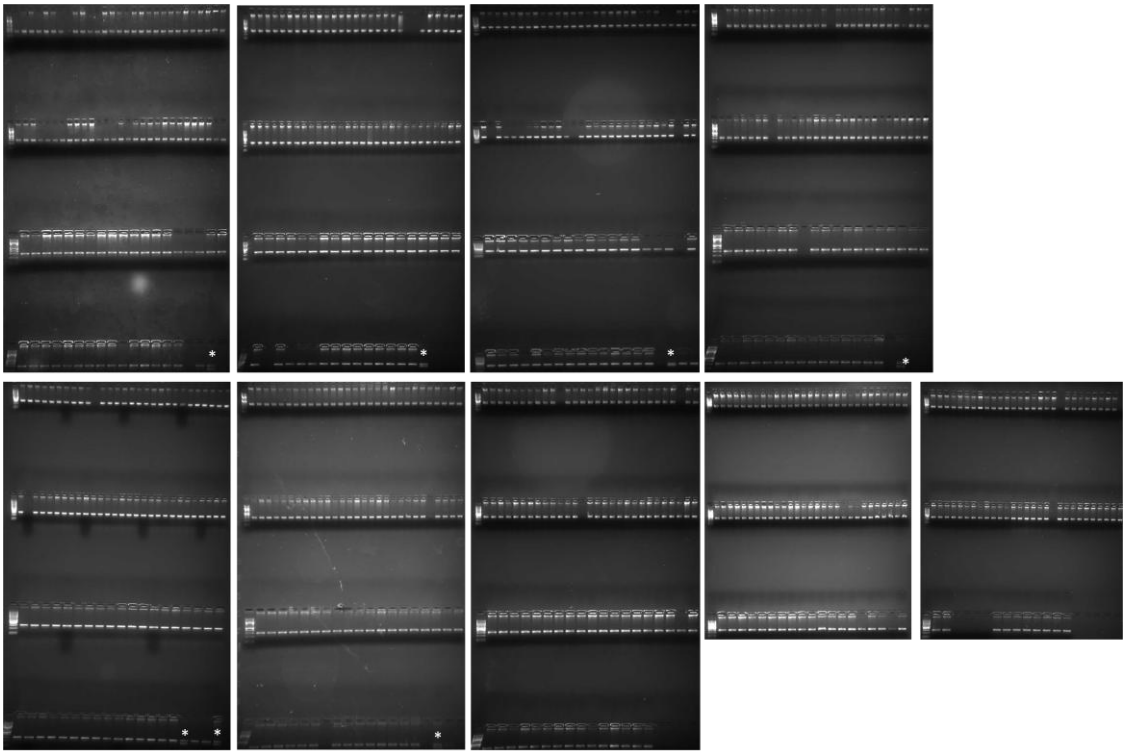


Figure 5.2-9 Diagnostic PCR in 778 additional Cypriot individuals

CFHR5 duplication was not demonstrated in any of these individuals. Positive controls were included with each plate and are marked by white asterisks.

This demonstrated that the mutation occurred on a haplotype which was common in the Cypriot population, consistent with the mutation having occurred within Cyprus. This result indicates that the CFHR5 mutation is not simply tagging a rare haplotype which was imported into Cyprus and could hypothetically contain a 'second mutation' responsible for the disease. The similarity of frequency of the disease-linked haplotype in Pedhoulas compared with Nisou does not provide evidence that the Troodos region is a genetic isolate which is distinct from the Mesaourian plain, consistent with a model in which bidirectional gene flow between the Troodos region and the rest of Cyprus has allowed spread of the mutation across the island.

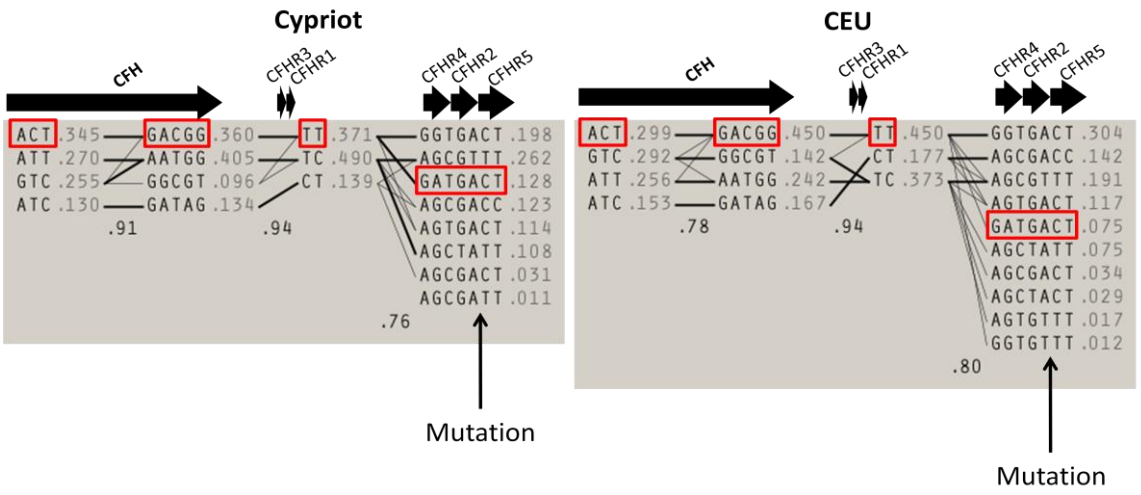


Figure 5.2-10 Haplotype blocks across the CFH/CFHR cluster

Common Haplotype blocks in the Cypriot and Caucasian populations across the CFH/CFHR gene cluster are shown. Individuals with the mutation all harboured the haplotype marked with a red box (which therefore lies on the green extended haplotype shown in Figure 5.2-7). This haplotype is more common in Cypriots than in the US Caucasian population (CEU).

5.2.4 Summary

CFHR5 nephropathy is a highly penetrant and common cause of renal disease in Cypriots (both in Cyprus and abroad) and detection of the CFHR5$^{12123\text{-}9}$ mutation is a robust clinical test. The disease is not confined to the Troodos region of the island and is likely to have existed for at least 2500 years. There is, as yet, no evidence to support the hypothesis that the mutation is prevalent in other populations, although a logical next step would be to test individuals with inherited C3GN or microscopic haematuria and renal disease in other Eastern Mediterranean countries. All the Cypriot individuals with C3GN tested so far have been shown to harbour the CFHR5$^{12123\text{-}9}$ mutation and there exists no evidence in favour of the existence of a putative 'second mutation' in a gene other than *CFHR5* but within the common haplotype which is responsible for the disease. This is issue will most likely only be resolved by identification of additional mutations in *CFHR5* which cause a similar phenotype.

5.3 Discussion

The findings presented in this and the previous chapters prove the existence of an autosomal dominantly inherited renal disease, characterised by C3 deposition in the glomerulus, which is endemic in Cyprus. The disease cosegregates with a heterozygous internal duplication of CFHR5 which is likely to cause dysregulation of complement at the glomerular surface, providing a mechanism which may explain the pathophysiology. These findings raise three important questions:

1) Why has CFHR5 nephropathy only now been recognised?

2) Why is CFHR5 nephropathy so common in Cyprus?

3) Why is CFHR5 nephropathy unrecognised in the rest of the world?

5.3.1 Why has CFHR5 nephropathy not been recognised before?

A crucial question is why, if it is such an important and common disease in Cyprus, has CFHR5 nephropathy only now been recognised? This is an important question to consider, not least because understanding the answer may allow determination of whether the disease is likely to be found in other populations. The answer is probably a combination of the following factors:

Firstly, provision of expert nephrology in Cyprus is a recent development, with a dedicated service offering reliable renal biopsy examination only being available in the last 25 years or so. Family history in many pedigrees reveals the existence of men in previous generations who died young (sometimes of kidney failure) who were never seen by a nephrologist. Retrospective review has revealed that several affected individuals (both in Cyprus and London) were diagnosed with IgA nephropathy on clinical grounds and never underwent renal biopsy. Furthermore, in at least one case where presentation was with classical microscopic and synpharyngitic macroscopic haematuria, the absence of mesangial immunostaining for IgA on

examination of the kidney biopsy was judged to be more likely a consequence of imperfect sample processing than owing to the presence of a disease entity not described in the published literature (Dr A Pierides, personal communication).

Secondly, renal impairment in CFHR5 nephropathy is rare in females, children and young adults. Most of the morbidity and mortality is borne by males over the age of 50, a high proportion of whom are hypertensive. This has led to under-appreciation of the familial nature of the condition. In addition, many clinicians (in both the UK and Cyprus) regard this population with 'diagnostic nihilism:' the widely-held (and not entirely unreasonable) view that chronic kidney disease in middle age (in the absence of nephrotic syndrome or features of systemic diseases such as vasculitis) is unlikely to be amenable to therapy other than blood pressure control, and therefore the value of obtaining a pathological diagnosis does not justify the risk associated with invasive investigations such as a kidney biopsy. In addition, many of the affected individuals only presented to a nephrologist at a time when they had advanced renal disease and small kidneys – a scenario in which renal biopsy carries an especially high risk and has a low diagnostic yield.

Thirdly, initial presentation of CFHR5 nephropathy is usually with isolated microscopic haematuria which, under current guidelines, is not investigated further (Cohen and Brown 2003; Halpin 2008). The lack of (or very slow) disease progression in many affected individuals has led to the under-appreciation of the number of affected people and the autosomal dominant nature of inheritance. In addition, the recent description of familial microscopic haematuria with progression to focal segmental glomerulosclerosis associated with thin glomerular basement membranes and mutations in *COL4A3* and *COL4A4* (also endemic in Cyprus) had led nephrologists on the island to believe that some of the families which turned out to have CFHR5 nephropathy (and manifested autosomal dominant microscopic haematuria and renal failure) were actually harbouring mutations in collagen-related genes. This illustrates

the importance of open-minded investigation of individuals and families using histological and molecular techniques.

Finally, the morphological categorisation of renal biopsy pathology is not always conducive to appreciation of the pathophysiology of the underlying disease. Almost all the patients with CFHR5 nephropathy who were biopsied were initially 'diagnosed' with primary mesangiocapillary glomerulonephritis type 1 (MCGN type 1) on the basis of the light microscopic findings. MCGN has a number of recognised causes (***Table 5.3-1***) and the remaining cases are regarded as 'primary MCGN type 1' (Jeanette 2008) which is said to carry a poor prognosis and is not amenable to treatment. Furthermore, IgA nephropathy and SLE both frequently produce the morphological appearances of MCGN type 1 on light microscopic examination, but the other clinicopathological features usually allow the clinician and pathologist to make the correct diagnosis: a biopsy report of 'MCGN type 1' in this context (while not being incorrect) is usually highly misleading. There is now growing evidence to suggest that 'primary MCGN type 1' is not a useful diagnostic entity at all, and cases could more usefully be regarded as 'MCGN type 1 of unknown cause.' The corollary is that 'MCGN type 1' is analogous to 'purpuric rash' in that it is a distinct clinical feature which has finite (if diverse) aetiologies, each defined and diagnosed by the presence or absence of other clinical and pathological features. In CFHR5 nephropathy, recent advances in the understanding of the biology of complement regulation, coupled with the appreciation of C3GN as a distinct subset of MCGN type 1 (usually caused by genetic or acquired defects in complement alternative pathway regulation), allowed recognition of the pathophysiological unity of the disease affecting families 1 and family 2, and hence identification of the mutation.

Type	Category	Example
Immune complex–mediated disease	Idiopathic MCGN or of unknown association	Mesangiocapillary glomerulonephritis type I
		Mesangiocapillary glomerulonephritis type II or dense deposit disease and PLD
		Mesangiocapillary glomerulonephritis type III
	Autoimmune diseases	Systemic lupus erythematosus (SLE)
		Sjögren syndrome
		Rheumatoid arthritis
		Inherited complement deficiencies, in particular, C2 deficiency
		Scleroderma
		Celiac disease
	Chronic infections	Viral - Hepatitis B, hepatitis C, and cryoglobulinemia type II
		Bacterial - Endocarditis, infected ventriculoatrial (or jugular) shunt, multiple visceral abscesses, leprosy
		Protozoal - Malaria, schistosomiasis
		Other infections - Mycoplasma
	Miscellaneous	Chronic liver disease (cirrhosis and alpha1-antitrypsin deficiency)
Chronic and recovered thrombotic microangiopathies		Healing phase of hemolytic uremic syndrome and/or thrombotic thrombocytopenic purpura
		Syndromes of circulating antiphospholipid (anticardiolipin) antibodies
		Radiation nephritis
		Nephropathy associated with bone marrow transplantation
		Sickle cell anemia and polycythemia
		Transplant glomerulopathy
Paraprotein deposition diseases		Glomerulonephropathies associated with cryoglobulinemia type I
		Waldenström macroglobulinemia
		Immunotactoid glomerulopathy
		Immunoglobulin light chain or heavy chain deposition diseases
		Fibrillary glomerulonephritis
		Monoclonal gammopathy of unknown significance
Malignant neoplasms		Lymphoma
		Leukemia
		Carcinoma

Table 5.3-1 Causes of mesangiocapillary glomerulonephritis
MCGN, or alternatively membranoproliferative glomerulonephritis, after (Jeanette 2008).

5.3.2 Why is CFHR5 nephropathy so common in Cyprus?

In an isolated population, mutations can become established (i.e. their frequency increase) in a population as a consequence of one of two effects. Either they can confer some phenotypic characteristic which results in increased ability to transmit the allele to the next generation (i.e. positive selection) or else the frequency can change as a consequence of genetic drift. Conventionally, positive selection in human Mendelian diseases is regarded as associated with autosomal recessive disorders, in which the survival advantage occurs in the heterozygous state (such as in sickle cell trait).

Genetic (or allelic) drift refers to changes in allele frequency caused by random fluctuations in transmission rate from one generation to the next. Owing to the factorial nature of the binomial probability distribution, these changes are much more pronounced in small populations compared with large ones[a] and since the expected allele transmission frequency depends on the frequency of the allele *only* in the preceding generation (i.e. it obeys the Markov property – see Chapter 1), over time rare alleles are more likely to be either lost or amplified in small populations compared with larger ones, in which their frequency is less likely to change. Since Cyprus is an island and therefore has been a relatively isolated population for much of its history, both genetic drift and selection could theoretically be responsible for the observed prevalence.

Many of the common disease-associated genetic variants are believed to confer a survival advantage on those who harbour them, at least in some environmental contexts. In particular, there is now good observational evidence that α-thalassaemia alleles, in both the heterozygous and homozygous states, and sickle cell trait confer protection against severe falciparum malaria in children (Williams, Mwangi et al. 2005; Williams, Wambua et al. 2005). In addition, a host of other abnormalities of haemoglobin and red cell structure have been documented in regions in which malaria is endemic and many have been reported to confer protection from the disease (reviewed in (Williams 2006)).

Malaria was endemic in Cyprus until its vector (the Anopheles mosquito) was eradicated in 1946 (Constantinou 1998), an achievement for which the island's then governor, Lord Winster, won a cask of wine (Hansard 1960). However, the genetic legacy of the disease remains evident by the high prevalence of haemoglobin variants on the island (for example the β-globin gene variants Hb Limassol (β8(A5)Lys→Asn) and Hb Nicosia (β17(A14)Lys→Gln) as well as at least 10

[a] Analogous to the large disparity in the probability of throwing 7 or more heads in 10 coin tosses ($p=0.172$) compared with throwing 70 or more heads in 100 coin tosses ($p<0.0001$) despite the same percentage deviation from the expected number.

other variants which are present with a combined frequency of 0.4% of Cypriots. However, although some of these variants have only been described in Cypriots, haplotype analysis has indicated that others were imported into Cyprus, probably in historical times: the commonest variant is the HbS allele which is predominantly found along the Northern coastal region of the island and is linked to the so-called 'Benin' haplotype. It is likely that the allele was transmitted from Benin along the trans-Saharan trade routes to North Africa and thence to the Eastern Mediterranean region where it has been identified in several populations, including Greeks, Syrians and Turks (Kyrri, Felekis et al. 2009). The historical record indicates that, between the 5th and 10th centuries AD, Northern Cyprus was frequently raided by pirates based in North Africa. At that time the island was under Byzantine rule and was guarded by a cosmopolitan army from all over the far-flung Byzantine Empire which included black Africans. It is therefore easy to postulate the introduction of genetic variants during this period but what is less clear is the extent to which these alleles have become established and maintained within Cyprus as a consequence of positive selection as opposed to genetic drift in a relatively isolated population.

5.3.2.1 Innate immunity and susceptibility to disease in childhood

While CFHR5 nephropathy is associated with significant morbidity in late adulthood, symptomatic renal disease has only been observed in individuals beyond reproductive age, implying that any negative effect on reproductive ability would be relatively minor. In considering how the *CFHR5* mutation might offer a selective advantage, one possibility is that could increase the likelihood of surviving childhood, during which period the major threat to life (especially in the pre-antibiotic era) arises from infectious diseases. The first months of life are recognised to pose a particularly high risk of death from infection and this is thought to result from immaturity of a range of innate and adaptive immune mechanisms, including bone marrow exhaustion, failure of neutrophil chemotaxis and microbial killing and relative complement deficiency (Wilson 1986). In addition, young children exposed to an invading microorganism for the first time will not have developed specific immunity to it and will be

dependent either on passive immunity (i.e. from maternal immunoglobulins) or innate defences.

Variation in a range of genes coding for components of the immune system exert their greatest effect in the young. It is well recognised that diseases such as common variant immunodeficiency and severe combined immunodeficiency (which are consequences of mutations causing severe defects of adaptive immune function) most commonly present with infectious diseases in infancy. In addition genetic variation in the innate immune system can also modify infection risk in children, with examples being mutations in Mannose Binding Lectin, which are over-represented in children hospitalised for infection compared with controls (Summerfield, Sumiya et al. 1997; Cedzynski, Szemraj et al. 2004); Complement Factor I deficiency, which is associated with increased susceptibility to infection (Amadei, Baracho et al. 2001); and Toll-like Receptor and Tumour-Necrosis Factor-α polymorphisms, which are associated with altered susceptibility to childhood *Helicobacter pylori* infection (Wilschanski, Schlesinger et al. 2007; Moura, Almeida et al. 2008).

Although most of the identified variants are associated with increased susceptibility to infection, recently, it has been shown that rare variants of the genes *TRIM5α* and *CCR5* are associated with *decreased* susceptibility to HIV (Reiche, Ehara Watanabe et al. 2008; Nakajima, Nakayama et al. 2009). Clearly, it is difficult to argue that HIV has exerted selection pressure on the populations in which these variants have been described, since the duration of the HIV pandemic has been so short – it appears more likely that the protective variants exist as (previously) neutral polymorphisms in the population. However, it may be that, in the absence of effective therapy, positive selection for these variants would occur in the future. Alternatively these variants may have been under positive selection in an ancestral population in which a similar retroviral pathogen was endemic.

In addition to variation in components the immune system (and as discussed above), disease-causing variants of haemoglobin and red cell structure and function have been shown to confer protection against malaria. Advances in pharmacology, public health and sanitation in most developed countries (including Cyprus) have decreased the burden of disease caused by infectious organisms and have shifted the balance from the beneficial and towards the detrimental effects of these genetic variants. Clearly, in an environment where malaria is either absent or effectively treated pharmacologically, there is little or no reproductive benefit conferred by thalassaemia or sickle cell traits: in clinical practice in London, for instance, the disease burden of sickle cell disease far exceeds that of malaria.

Together these findings demonstrate that mutations (even those responsible for disease) can confer protection against infectious diseases in some circumstances. This provides at least a biological precedent for the hypothesis that CFHR5 nephropathy, while detrimental in adults in modern societies, may confer an advantage to children exposed to infectious diseases which was significant in the pre-modern era. Support for this hypothesis (in the modern era with low childhood mortality from infectious disease) could be obtained by an *in vitro* demonstration of increased microbial killing of serum harbouring CFHR5$^{12123\text{-}9}$ compared with wild type CHFR5.

5.3.3 CFHR5 nephropathy in the rest of the world

Clearly, the fact that CFHR5 nephropathy is not a recognised disease entity in other communities could have two explanations: either it only exists in the Cypriot population or else it exists in other populations but is not recognised as a disease entity there owing to lack of appropriate investigation (see above). The results presented earlier in this chapter have failed to identify direct genetic evidence that the CFHR5$^{12123\text{-}9}$ mutation was imported into Cyprus from a separate parent population (since the haplotype on which the mutation exists is common in the Cypriot population), and it remains unclear whether or not the disease existed prior to the arrival of the ancestors of the majority of the current inhabitants of the island. If it

did, then it would be reasonable to hypothesize that other populations in the Eastern Mediterranean region would be at risk of harbouring the disease. The prevalence of the allele in such a 'cousin' population would depend on considerations of genetic drift, selection (see above) and admixture with neighbouring or migrating populations on the continental mainland.

5.3.3.1 Cypriot origins

In order to determine whether a population outside Cyprus is likely to harbour individuals with the mutation it is necessary to understand the genetic heritage of modern Cypriots. Owing to ethnic conflict in the late 20th century, the question of Cypriot origins has significant political implications and is an area fraught with claim and counter-claim. Cultural considerations have led to the popular perception that 'Greek Cypriots' are closely related to people from Greece, and 'Turkish Cypriots' are closely related to people from Turkey. While this may be true on linguistic, cultural and religious grounds, consideration of archaeological, historical and genetic evidence suggests that Greeks, Turks and Cypriots are each descended from very different ancestral populations.

Archaeological record in Cyprus

The archaeological record indicates that the first evidence for human habitation in Cyprus dates to the early Neolithic period (approximately 10,000 to 6000 BC). These inhabitants are thought to have introduced animals, such as dogs, sheep, goats, deer and foxes which together probably led to the extinction of native pygmy hippopotamus and dwarf elephant. No ceramic artefacts have been identified in the remains left by these early inhabitants and the type of circular architectural ruins they produced is similar to the pre-pottery Neolithic B culture which is thought to have its origins in Northeastern Anatolia or the Levant (the Mediterranean coast of Asia). No remains at all have been found dating from 6000 BC until around 4500 BC, consistent with extinction of this original community and coinciding with the '8.2 kiloyear event' which was a cooling in the climate lasting approximately 4 centuries from 6200 BC.

Whether the inhabitants actually died out during this period or simply ceased producing buildings or burial sites is not known.

The architectural record resumes with the Neolithic II (or ceramic Neolithic) period (4500 to 3800 BC) which was destroyed following an earthquake and was followed by the Chalcolithic period which lasted from 3800 to 2300 BC and was characterised by more variation in the architectural remains left behind and the first signs of metalwork. Remains from this period are mostly copper, but traces of tin imply contact with Anatolia (what is now Asiatic Turkey) during this period.

The Bronze Age began in Cyprus in around 2400 BC and the distinctive Philia Culture architectural remains attests to import of this technology from Anatolia. Regions of Cyprus, in particular the foothills of the Troodos mountains, are rich in copper and evidence of copper export from Cyprus and significant international trade dates from this time through to the late Bronze Age (Knapp, Kassianidou et al. 2001). The late Bronze Age also saw the appearance of Cypriot pottery as far afield as Palestine and Crete, and the first architectural features characteristic of those found on the Greek mainland (namely Cyclopean walls) appeared in Cyprus. While this is compatible with Greek settlement, other archaeological remains demonstrate that links with other regions (including Egypt) were present in this period. Intriguingly, one of the earliest ever representations of the human kidney was a bronze model dated to the 13th century BC which was found with an inscription in the Cyprominoic script at Kition in Cyprus (see ***Figure 5.3-1***). Whether this represents an early appreciation of the importance of kidney disease in Cyprus remains to be determined.

Figure 5.3-1 One of the earliest known human representations of the kidney
This bronze sculpture was found in Kition, in Cyprus and is thought to date from the 13th Century BC, after (Marketos, Eftychiadis et al. 1993).

History of Cyprus

Cypriot mythology links the early Iron Age with the foundation of numerous towns by Greek heroes following the Trojan war (examples being Salamis which is traditionally believed to have been founded by Teucer, brother of Aias and cousin of Hector; and Paphos which was purportedly founded by Agapenor, King of the Arcadians). Physical evidence of Greek influence in Cyprus comes from Greek inscriptions (albeit written in the Cypriot syllabic script) found on a tomb dating from the 11th century BC. Further evidence of Greek cultural influence (for instance cremation of the dead) is seen in the following centuries, alongside the appearance of Phoenician colonisation.

The historical record shows that Cyprus was under Assyrian rule in the 8th Century BC. It was then independent for approximately 100 years but came under the political control of the Egyptians and then the Persians in the 6th Century BC. Alexander the Great gained control of the island which then passed to the Ptolemaic kingdom and remained firmly within the Hellenic world (situated as it is near the important trade route between Alexandria and Athens) for most of the period until Roman rule started on the island in 58 BC.

Following the division of the Roman Empire in 285 AD, Cyprus lay within the Eastern Roman Empire and was ruled from Byzantium until it was invaded by the Arabs in 688. For the ensuing 3 centuries the island was governed jointly by the Arabs and the Byzantines until it was reconquered by the Byzantines who controlled Cyprus until the 12th Century when it fell under Crusader control, initially being ruled by Richard the Lionheart of England and subsequently by the Knights Templar and then the Franks who established an independent Kingdom of Cyprus (under Guy of Lusignan) with the Latin language and Church. The Maronites arrived in Cyprus during this period and a few communities persist in the North of the island. The Lusignan dynasty ended when the last member sold the island to the Venetians in 1489.

Cyprus came under Ottoman rule in 1571 and a significant proportion of the population died or fled, both as a result of bloody uprisings and also the plague which ravaged the country between then and 1700. Cyprus was occupied by the British in 1878 owing to its strategic position near the Suez canal and was annexed in 1914. The island achieved independence from Britain in 1960 and was partitioned in 1974 into the Republic of Cyprus (Southern Cyprus) and Turkish administered Northern Cyprus, which declared itself the Turkish Republic of Northern Cyprus in 1983 and is recognised as such only by Turkey.

Phylogeography of Cyprus

While providing significant insight into the flow of language, religion and culture into and out of Cyprus, the archaeological and historical records are less good at revealing the bulk

movements of people: political and cultural influence in one country from another only rarely results in replacement of the people living there – more commonly the inhabitants simply adopt (and subsequently adapt) the language and culture of the ruling nation. To address the question of the *genetic* origins of modern Cypriots, the genetic record is a potentially important source of information. While much work remains to be done in this area, it is clear that mitochondrial DNA differs significantly between modern inhabitants of Cyprus and Northern Greece (Irwin, Saunier et al. 2008) and principal component analysis of 500,000 autosomal genetic markers in Europeans from multiple countries indicates that, in general, genetic similarity is very tightly correlated with geographical proximity in European populations and that Cyprus is 'distant' from both Greece and Turkey (***Figure 5.3-2***, after (Novembre, Johnson et al. 2008)). Specific analysis of mitochondrial DNA variants in modern Cypriots, Cretans and people from the Levant has indicated the presence of variants in Cyprus found in the neighbouring regions (while certain variants which are common in the Levant are absent in people from Crete) (GENEMILL 2003). Importantly, Eastern Eurasian variants were completely absent in Cyprus although they are found in populations speaking Altaic languages (i.e. an arc between Turkey and Northeast Asia/Japan). This suggests that immigration into Cyprus occurred *prior* to the spread of East Eurasian populations westward.

Together, these findings suggest that the genetically closest relatives of modern Cypriots are likely to reside in the Levant and the ancestral population was likely to have lived there several thousands of years ago – probably in Neolithic times. In view of the estimate for the minimum age of the mutation, it remains unclear whether CFHR5$^{12123\text{-}9}$ arose within Cyprus in prehistoric times or was imported into the island by the ancestral population from the Levant. It does, however, appear less likely that the mutation was imported by any of the waves of settlement or occupation of the island which occurred in historical times. It is also likely that affected individuals from Northern Cyprus will be identified if enough individuals from this

population are tested, since the mutation is likely to antedate the linguistic and cultural division of the island.

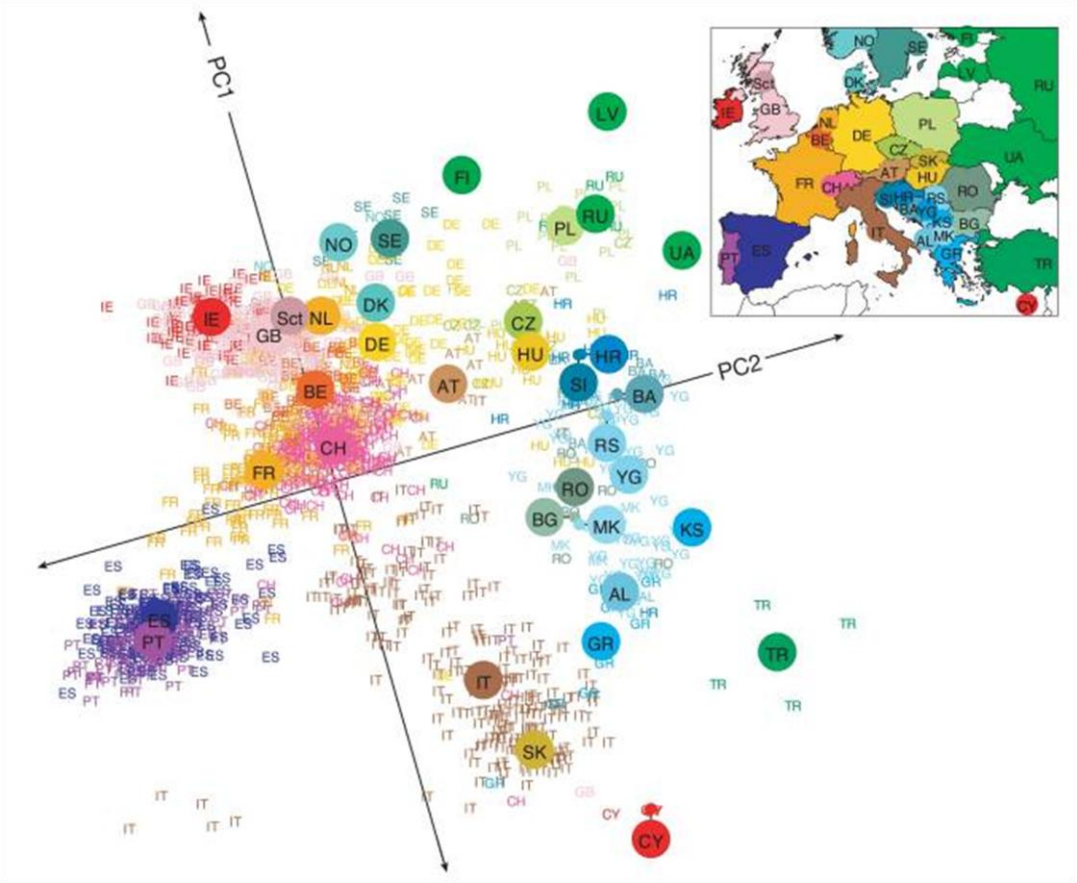


Figure 5.3-2 Genes mirror geography in Europe

Principal component analysis of autosomal genetic variants among different European populations. There is a clear correlation with geographical location and Cyprus is an outlier, being significantly different from both Greece and Turkey. After (Novembre, Johnson et al. 2008).

Appendix A: Conflicting Homozygosity Analysis

6.1 Conflicting Homozygosity analysis – theory

Since current SNP genotyping technology yields diploid genotypes rather than phased bi-allelic data, the most likely haplotypes have to be estimated computationally from these data, and different algorithms have been employed to do this (for instance the maximum likelihood estimation approach used by GENEHUNTER (Kruglyak, Daly et al. 1996)). However, whatever the algorithm used, the problem of assigning haplotypes has been termed NP-hard (non-deterministic polynomial-time hard) because verification of the 'correct' solution is computationally as hard as generating that solution[a] (Gusfield 2001). In this context, increasing the density of SNPs genotyped increases the computational demands and in any case still relies on whether sufficient informative individuals from the family have been genotyped. In practice, linkage studies are performed on subsets of genotyped SNPs (for instance the program PLINK removes SNPs which are in LD with each other – reducing 300,000 SNP genotypes to approximately 10,000 SNPs which are then analysed). Running parametric linkage analyses on untrimmed dense SNP marker datasets is impractical, even with modern supercomputers. In addition, the very large number of possible inheritance vectors, each with a correspondingly minuscule likelihood estimate, potentially provokes computational malfunction as a result of rounding errors owing to the need for very small floating point numbers.

The conflicting homozygosity (CH) approach, rather than estimating the haplotypes and then asking whether they are co-transmitted with the trait, simply asks the question: "Is the genotype data at each SNP among a set of affected individuals compatible with identical-by-descent (IBD) inheritance?" If one or more individual in the set is homozygous for the minor allele and one or more individual is homozygous for the major allele of a SNP this excludes IBD inheritance of a haplotype which spans that SNP. Clearly, if genotypes from a small number of

[a] An example of an NP-hard problem is the travelling salesman problem: "Which is the shortest route which passes through all specified points exactly once?" This contrasts with the NP-complete problem: "Is there a route passing through all points exactly once of length less than 10 miles?" which is easy to verify for any proposed solution.

individuals are compared, most SNPs considered in isolation will be non-informative since, even if there is no IBD inheritance most individuals will be homozygous for the major allele or heterozygous. However, if a dense enough SNP array is used, such that an IBD haplotype contains a large number of SNPs, the chance of none of them demonstrating CH purely by chance (i.e. in the absence of IBD inheritance of one or two alleles from a common ancestor) begins to fall. In this situation, identifying large (i.e. >1 cM) regions of the genome for which there is no CH in multiple individuals implies co-inheritance of the same haplotype from a single common ancestor.

The Illumina HumanCytoSNP-12 panel consists of 300,000 SNPs across the human genome, which consists of approximately 3,000,000,000 base pairs spanning approximately 3,000 cM. The average spacing of SNPs is therefore 100 SNPs per cM. Although in an outbred population the likelihood of each allele of a given SNP may be dependent on the allele of neighbouring SNPs *on the same chromosome* (i.e. nearby SNPs may be in linkage disequilibrium), the probability of homozygosity is independent of this (since the allele of any SNP on the maternally inherited chromosome is independent of the allele of the paternally inherited SNP). Therefore, the probability of conflicting homozygosity of a given SNP in two individuals (in the absence of shared ancestry at the locus in question) is dependent only on the allele frequency of the SNP.

Consider a SNP with alleles A and B occurring at frequencies of p and (1-p). In order for conflicting homozygosity to occur at this locus in individuals X and Y would require genotypes of either AA in person X AND BB in person Y OR BB in person X AND AA in person Y. This probability of conflicting homozygosity occurring in any 2 individuals, *P(CH),* is therefore given by:

$$P(CH) = 2p^2(1-p)^2$$

And the chance of observing no CH in *n* independent consecutive SNPs is given by:

$$P(no\ CH) = \big(1 - P(CH)\big)^n$$

Since the average minor allele frequency in the Illumina HumanCytoSNP-12 panel is 0.22 (Illumina 2010) the probability of observing conflicting homozygosity for a given SNP in 2 individuals (who do not share ancestry at this locus) is 2 x 0.22^2 x $(1\text{-}0.22)^2$ = 0.059, corresponding to an expected value (from the binomial theorem) of 16.98 SNPs for each episode of CH, assuming independent inheritance of neighbouring SNPs.

While the likelihood of observing homozygosity (and hence CH) in a given SNP is only dependent on the allele frequency of that SNP, neighbouring SNPs along a chromosome are not actually independently inherited – rather, in the general (unrelated) population the alleles occur in haplotype blocks (or LD blocks). Of considerable importance is the size of these blocks which, in Northern Europeans typically extend 60 kbp, with substantial variation across the genome (Reich, Cargill et al. 2001). Given the size of the human genome, it can be estimated that a 300,000 SNP panel is likely to produce, on average, 6 SNPs per 60 kbp block.

The extended haplotypes which are inherited identical-by-descent from a recent common ancestor are several orders of magnitude larger than these LD blocks, starting at an average of 50 cM (approximately 50,000 kbp) in size and (on average) halving with each generation.

For a single SNP, homozygosity can arise either because of inheritance of 2 copies of the same haplotype block (haplotype homozygosity), or alternatively because of inheritance of 2 different haplotypes which both happen to have the same allele at the SNP in question. Clearly, in the former case, the neighbouring SNPs within the block will also be homozygous, whereas in the latter case the probability of homozygosity of neighbouring SNPs is independent. As the haplotype diversity increases, the likelihood of haplotype homozygosity decreases (since haplotype frequency must be equal to or less than allele frequency at each SNP). In any group

of individuals, co-inheritance of the same haplotype block (as a function of the frequency of that haplotype block in the population, rather than because of a recent common ancestor) will result in a run of no CH the length of that block. This defines the lower limit of sensitivity of this approach, since completely unrelated people will therefore share runs of no CH as a consequence of shared haplotype blocks, as well as by chance. However, since the size of these blocks (in numbers of SNPs genotyped) is small compared with the expected number of SNPs between episodes of CH in non-IBD regions (~17 comparing 2 individuals), detection of haplotype blocks will not be possible with this density of SNP panel. Equally, the contribution that LD blocks make to the likelihood of observing run lengths will be small for runs spanning regions of the genome of significantly greater than the average block sizes (i.e. for runs >20 SNPs).

In the absence of inheritance from a common ancestor, the likelihood of observing n consecutive SNPs for which there is no conflicting homozygosity in 2 individuals genotyped with the Illumina HumanCytoSNP-12 panel is $(1\text{-}P(CH))^n = (1\text{-}0.059)^n$. It follows that, for any 2 individuals, observing a run of 50 consecutive SNPs (spanning, on average, 500 kbp) from the Illumina HumanCytoSNP-12 panel without encountering conflicting homozygosity is ~20 times more likely to occur as a consequence of shared inheritance of a haplotype at that locus (in which case identical-by-descent inheritance of one haplotype renders conflicting homozygosity impossible) than it is to occur by chance.

Clearly, when additional individuals are compared at the same locus, the absence of conflicting homozygosity across a region becomes even less likely to occur by chance, reducing according to a binomial distribution. For each SNP, the likelihood of observing no CH in N people who do not share ancestry at a given locus is $P(\text{no CH in N people}) = (1\text{-}P(CH))^N$

Therefore, L(no CH) the likelihood of observing no episodes of conflicting homozygosity across *n* SNPs in *N* people (in the absence of shared inheritance at the locus in question) is given by:

$$L(no\ CH) = \Big(\big(1 - P(CH)\big)^N\Big)^n$$

$$= \big(1 - P(CH)\big)^{Nn}$$

6.1.1 Effect of genotyping errors

Mendelian error checking (using PEDCHECK) demonstrated that the genotyping error rate was very low (<0.1% genotypes). Genotyping error could result in failure to detect homozygosity when it was actually present potentially leading to failure to detect conflicting homozygosity. However, multiple consecutive errors of this type would be needed to produce a significant increase in the observed run length of SNPs for which there is no conflicting homozygosity. This is illustrated in ***Figure 6.1-1***, in which the number of consecutive SNPs for which there is no CH is plotted against position along the chromosome. In the areas where there is no IBD inheritance (and hence CH occurs according to the distribution outlined above) the run length varies (randomly) between 0 and ~20 SNPs. Where there is IBD inheritance (and hence CH is impossible) the run length increases.

The other potentially important consequence of genotyping error (or non-Mendelian inheritance) would be the introduction of conflicting homozygosity in areas which are actually inherited IBD by two or more individuals. However, unless there was clustering of multiple types of this error, it would result in a single episode of CH (as a consequence of error) flanked by regions in which there is no CH (as a consequence of IBD inheritance) – leading to a double peak on a graph such as the one shown in ***Figure 6.1-1***.

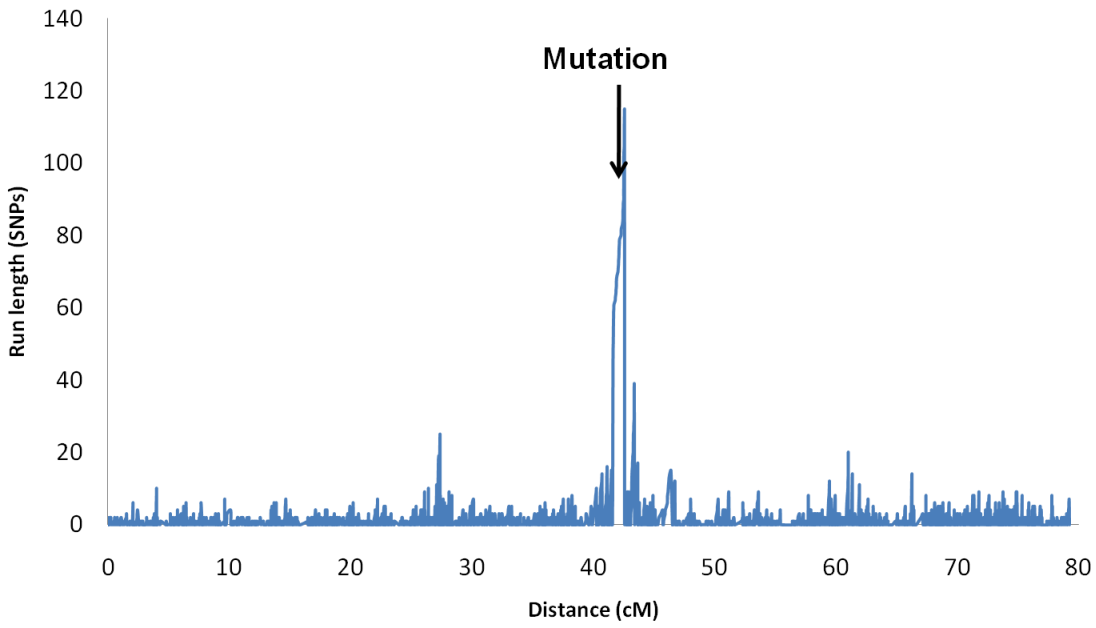


Figure 6.1-1 Conflicting Homozygosity analysis yields clear IBD signal
The run length (i.e. number of consecutive SNPs) for which there is no conflicting homozygosity (CH) in 17 ostensibly unrelated Cypriot families is plotted against position across an 80cM region of chromosome 1. All families are known to share a mutation

6.2 Conflicting Homozygosity analysis – empirical validation

This analysis was performed for the 6123 SNPs located in the 80cM spanning CFHR5 in 17 index cases, all harbouring the CFHR5$^{12123\text{-}9}$ mutation, each from a different family. This produced an average spacing of 0.0129 cM between each SNP across this region (or 78 SNPs per cM) and the result of CH analysis is represented graphically in ***Figure 6.2-1*** in which the cumulative number of occurrences of conflicting homozygosity, N(CH), in each of 16 individuals compared with the 17th individual who harboured the ancestral haplotype across CFHR5 (i.e. the longest run of no conflicting homozygosity with any other family) is plotted against distance from the mutation in number of SNPs. This graph demonstrates that there is an area of no conflicting homozygosity flanking the mutation in all the families (i.e. consistent with IBD inheritance) and in each family this zone of N(CH) = 0 is flanked by genomic regions where N(CH) increases linearly with distance. A linear regression line plotted through the points for

which N(CH) > 0 (i.e. where there is no identical by descent inheritance) has a slope ΔN(CH)/(number of SNPS) which is the reciprocal of the number of SNPs per episode of conflicting homozygosity. For all 16 families tested, the mean of the regression coefficients was 0.056475, corresponding to an average of 17.7 SNPs per episode of conflicting homozygosity. The mean minor allele frequency for all the SNPs in all the individuals included in this analysis (in these 17 individuals) was 0.211439. It is notable that:

$$\frac{1}{2 \times 0.2114392^2 \times (1 - 0.2114392)^2} = \frac{1}{0.055599} = 17.98578$$

Which is the expected number of SNPs per episode of CH across this region of the genome and is in good agreement with the observed distribution of CH. This implies that, as predicted above, the effect of haplotype blocks in individuals from this population is small enough not to cause a significant increase in average non-CH run length.

The empirical data therefore supports the hypothesis that occurrences of conflicting homozygosity (in areas with no recent IBD inheritance) obey this binomial probability distribution and hence that the absence of conflicting homozygosity over large enough distances is evidence of identical by descent inheritance. This result also implies that the effect of the population LD structure is small, as predicted by the small size of LD blocks (compared with the size of a recent IBD haplotype).

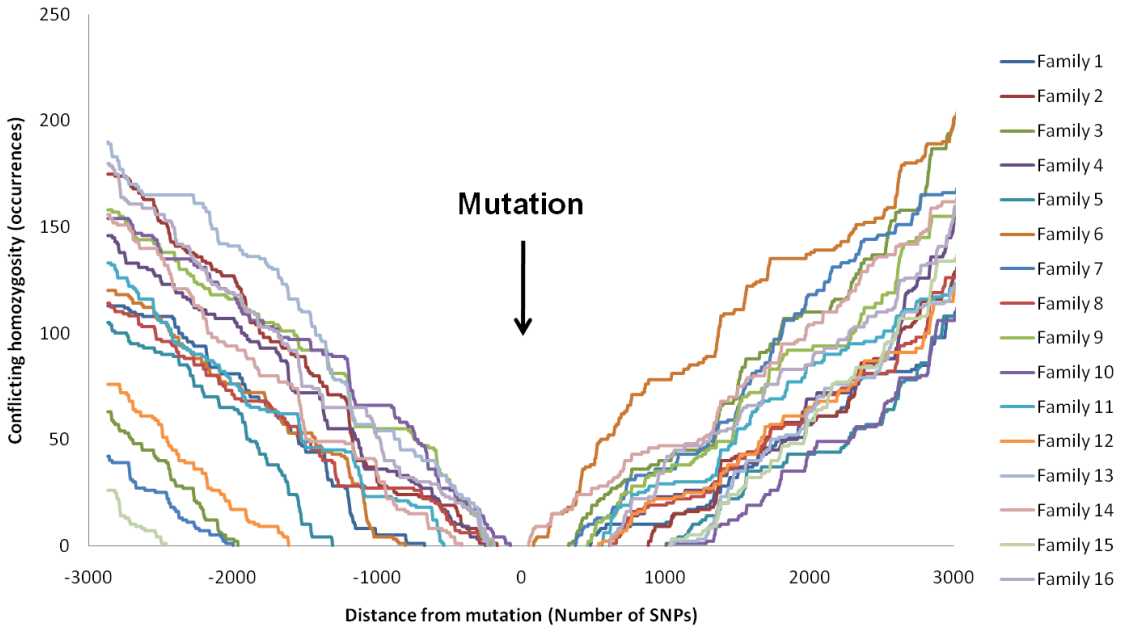


Figure 6.2-1 Likelihood of CH is constant in non-IBD regions

Cumulative number of episodes of Conflicting Homozygosity (CH) across an 80 cM region of chromosome 1 spanning CFHR5. 17 ostensibly unrelated Cypriot families known to share a mutation in CFHR5 were compared. A common haplotype in which no CH occurs spans the locus in all families. Outside the shared haplotype (identical by descent, IBD, region), dy/dx is constant and approximately equal in all families indicating a similar, fixed probability of observing CH in any SNP in regions not inherited identically by descent.

The frequency of run length of consecutive non-CH SNPs occurring in 17 families across the whole genome was measured and the result for run lengths >40 SNPs) is shown in ***Figure 6.2-2***. The box indicates the single run of 115 consecutive SNPs which occurred in the one region of the genome known to be shared IBD by all 17 families (i.e. the locus containing CFHR5). There were 2 runs of ~90 SNPs in regions not known to be inherited IBD by all the families (i.e. likely to be observed by chance) and an exponentially rising number of loci with smaller run lengths of non-CH SNPs. Even if the *a priori* low probability of observing a run of this length is ignored, simply considering that the human genome spans in excess of 3000 cM the likelihood of observing no CH across only the 1 cM region spanning the mutation purely by chance can is less than 1:3000. To see a run length of this size occurring *only* at the locus inherited IBD by all the individuals is therefore extremely unlikely to have occurred except as a consequence of IBD inheritance. Repeating the analysis with the inclusion of a single unaffected relative in the

analysis completely removed the signal from the *CFHR5* locus. This experiment therefore demonstrated that CH can be used to exclude IBD inheritance at a locus in ostensibly unrelated families, and its absence is evidence for IBD inheritance.

A logical next step would be to use this type of analysis within a family in order to identify any regions of the genome inherited IBD by affected members of a kindred.

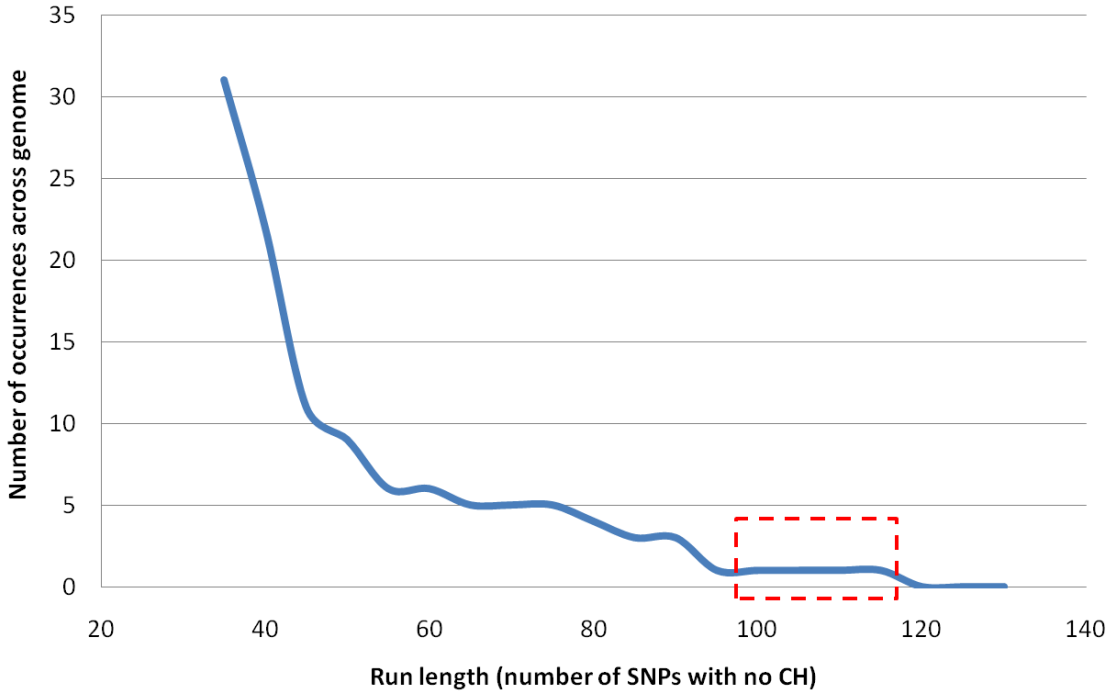


Figure 6.2-2 IBD signal is rare across the genome
In 300,000 SNPs genotyped in a single member of 17 unrelated Cypriot families the only run length >100 SNPs in the entire genome (red box) occurred in the region known to be inherited IBD in all the families since it spanned the shared CFHR5 mutation.

Appendix B: Materials and Methods

7.1 Ethical Approval

All research involving human participants was performed with informed consent and was approved by the Hammersmith and Queen Charlotte's and Chelsea Research Ethics Committee (06/Q0406/151) according to the principles of the Declaration of Helsinki. The information sheet and consent forms used are found in Appendix C.

7.2 DNA Extraction

Genomic DNA was extracted from either EDTA-preserved whole blood (QiaAMP, Qiagen, Germany) or saliva (Oragene, DNA Genotek, Ontario, Canada) according to the manufacturer's instructions.

7.3 Single Nucleotide Polymorphism (SNP) Genotyping

SNP genotyping for linkage studies was performed at the Genome Centre, Queen Mary University London using the Illumina Linkage IV 6008 SNP panel according to the manufacturer's instructions. Additional SNP genotyping was performed at the Genome Centre, University College London, using the Illumina Cyto-SNP12 300,000 SNP panel according to the manufacturer's instructions.

7.4 Linkage analysis

Analyses was performed on a Hewlett-Packard desktop PC with a 2.66 GHz Intel Core Duo CPU and 2.0 GB RAM running either Microsoft DOS or Microsoft Windows – either XP Professional (32-bit) or Vista (64-bit). Tab-delimited text files containing SNP genotypes in AB/00 format were generated using JUJUEDIT (Jujusoft, USA) and EXCEL2007 (Microsoft, USA). Pedigree files were written using NOTEPAD++ (Sourceforge). The data were first interrogated for missing data (using EXCEL) and Mendelian errors (using PEDCHECK). An individual in whom

>100 Mendelian errors were detected was re-genotyped from a fresh DNA sample. Repeat analysis indicated <10 Mendelian errors involving this person. SNPs with Mendelian errors were omitted from further analyses. Data were analysed in duplicate firstly using PEDCHECK and MERLIN under DOS and secondly using PEDCHECK, GENEHUNTER and MERLIN running behind the EasyLINKAGE graphical user interface under Windows. Parametric linkage analysis was performed using a dominant model (penetrance vectors: wt/wt: 0; wt/mt: 1; mt/mt 1) and a codominant allele frequency algorithm. Non-informative subjects (unaffected/unknown individuals with genotyped parents and no affected descendants) were discarded from kindreds with >15 meioses.

Analyses were performed using sets of 100 markers and repeated using sets of 71 markers (to prevent loss of accurate haplotype generation at section boundaries).

7.5 Haplotype analyses

Haplotypes within each family were estimated using GENEHUNTER and used as input files for HAPLOPAINTER. The output of HAPLOPAINTER was used to compare directly predicted haplotypes at linked loci in affected members of all families. Haplotype analysis for estimation of identical by descent inheritance in unrelated individuals was performed on genotype data from 300,000 SNPs typed across the genome using the Illumina CytoSNP-12 panel. The average distance between each SNP is approximately 0.01 cM (i.e. there are approximately 100 SNPs per cM). SNP alleles were assigned numbers 1 and 2 (with allele A = 1 and allele B = 2) with possible genotypes therefore being 11, 12 or 22. At each SNP, numerical genotype for each individual was compared with the all other individuals using EXCEL. Numerical genotypes differing by exactly 11 indicated conflicting homozygosity and excluded identical by descent inheritance of a haplotype across this locus (see Appendix A).

7.6 PCR and primers

Unless otherwise stated, all PCR reactions using genomic DNA were performed with 1 unit Platinum Taq (Invitrogen, UK) according to the manufacturer's instructions with 50-200 ng template DNA; 200 µM dNTPs; 2.5 mM $MgSO_4$; and 400 nM primers. Sequencing primers were designed using PRIMER3 (http://primer3.sourceforge.net/) and blasted against the genome to ensure a single predicted product. For specific amplification of regions of high genomic complexity such as the *Complement Factor H/CFH-Related* gene cluster, the whole genomic region was loaded into SEQUENCE SEARCHER (Viral Bioinformatics) and a fuzzy search allowing 5 mismatches performed using each primer sequence. Primers where rejected if there was more than one match to the genomic region using this search strategy unless the 3'-nucleotide was one of the mismatches. Primers used for amplification of the exons and bordering intronic regions of the candidate genes sequenced are shown in ***Table 7.6-2***.

Amplification was performed with the following program:

Step	Temperature	Time
1	95°C	2 min
2	95°C	30 s
3	58°C	30 s
4	68°C	30 s

Table 7.6-1 PCR program for sequencing reactions
Steps 2 - 4 were repeated 34 times

Gene	Exon	Forward (5' → 3')	Reverse (5' → 3')
STK11	1	GGAAGTCGGAACACAAGGAA	GGGAGGAGAGAAGGAAGGAA
HIF2A	9	TGCATCTAGGGGAGCAGAAT	GCCAGGGAGGCCTGTTATAG
HIF2A	12	TTGAGCAGCACTGTGAAACA	ACATGGCTTGAGGTGATTCC
HIF2A	16	CAGACACCACTGAAGGAGCA	TGGGTTCCCCATAGGATACA
CFHR5	1	CAGTCCCATTTCTGATTGTTCCA	CCCCTTCAAATTATCCTCAGC
CFHR5	2	GGCATTTAAGCTGAATGAAAAA	TCCGGCACATCCTTCTCTAT
CFHR5	3	TGATGTCAGTTTTCAAAGTTTTCC	ATAATTAGCAAAACTGAGAGAGTGGT
CFHR5	4	CACATTAAATTTGTTTCTGCAATGA	TCACCTTCTTGTTTCATCACTTCT
CFHR5	5	CCATTTAAGCATTATTTATGGTTTC	TGCAAAGTAATAGTAACTGTCCTGTTT
CFHR5	6	AAATATTTTCAGAGTAAGCACTCATTT	ACAATCCCAATCAAAATGATAAA
CFHR5	7	TGCAGATATTTTATTGACATAATTGTT	TCTTGTAAAGAAGCAACAAGATCAAC
CFHR5	8	CCATTTTCCTGAAACACTACCC	TTGGGGTACAGTGCAACAGA
CFHR5	9	AATTATTTGAATTTCCAGACACCTT	GGGTTATTCTATGAAATTAGTCCAAAA
CFHR5	10	AGCCCATACACAGTGCTAAGAA	TGAGTTAAACAACACCTGAAAGAA
HIF2-ARMS	Inner	ACTGGCACCCTATATCCCCATGGCCA	CAGATGGGGCTTAGCTGGAAGTCTTCACC
HIF2-ARMS	Outer	AGTGCTTGAGATGAATGGCTCTGCAGGA	TTGCTCTCCAGCTGCTGCTGAAACTTGT
CFHR1	1	GCACAACCCTCCATGAACTT	GCATACACACAAAAACCGATAAA
CFHR1	2	TTCTTGAAGACAATGCAGGAGA	GGACTACATCTCCCATC
CFHR1	3	TGCAGTTGTACTTTTTCTTTGC	CCACTTGGACACAATTGGAA
CFHR1	4	TGATTTGCTACTCAAAATGAACACT	CCTGCTATACTCCCCCAAAA
CFHR1	5	TGCTGTGTTTGTATTTGCCTTA	TTTTTCCAGCCACGTGAATA
CFHR1	6	TTGAAAACCTGAAAGTCTATGAAGAT	TCTTGCCACAAACATTTTATGAAC

Table 7.6-2 PCR primers for amplification of genomic DNA

Amplification Refractory Mutagenesis System (ARMS) PCR primers (***Table 7.6-2***) were designed according to the method described previously (Ye, Dhillon et al. 2001). 400 nM inner primers were used but outer primers were present at a concentration of 40 nM. Amplification was performed using the following touchdown PCR program ***Table 7.6-3***:

Step	Temperature	Time
1	95°C	2 min
2	95°C	1 min
3	72 - 55°C	30 s
4	72°C	30 s
5	95°C	1 min
6	55°C	30 s
7	72°C	30 s

Table 7.6-3 PCR program for ARMS-PCR
Steps 2 – 4 were repeated 12 times with decreasing annealing temperature; 5 – 7 were then repeated 25 times

Specific amplification of the CFHR5 internal duplication used primers TGGAAGCCTGTGGTATAAATGA-3' and 5'-TCCGGCACATCCTTCTCTAT -3' amplified with the following program ***Table 7.6-4***:

Step	Temperature	Time
1	95°C	2 min
2	95°C	30 s
3	58°C	7 min
4	68°C	30 s

Table 7.6-4 PCR program for amplification of duplication
Steps 2 - 4 were repeated 34 times

The 3 primer diagnostic PCR which specifically amplified the mutant CFHR5 allele was performed with the primers 5'-GATTCCATTTGTCAAATATTG-3', 5'-TCTTCTCCAAAACTATCTAATGTCAA-3' and 5'-TTTGAATGCTGTTTTAGCTCG-3' and amplified with the following program ***Table 7.6-5***:

Step	Temperature	Time
1	95°C	2 min
2	95°C	30 s
3	58°C	30 s
4	68°C	30 s

Table 7.6-5 PCR program for 3-primer diagnostic PCR
Steps 2 - 4 were repeated 34 times

7.7 MLPA probes

Probes used in multiplex ligation dependent probe amplification (MLPA) assay are shown in

Table 7.7-1:

Product size	Gene	Exon/Intron	Probe sequence
256	KCNT2	Exon 3	TTACCATTCA-TTTCCTTGTG
399	KCNT2	Exon 1	GTTTCGAGAT-TTGCTGCTAG
202	CFH	Exon 1	TGCTACACAA-ATAGCCCATA
142	CFH	Exon 2	GGTCTGACCA-AACATATCCA
179	CFH	Exon 3	TCCTTTTGGT-ACTTTTACCC
419	CFH	Exon 4	ATTACCGTGA-ATGTGACACA
337	CFH	Exon 6	AAAGAGGAGA-TGCTGTATGC
238	CFH	Exon 9	AATCAAAATC-ATGGAAGAAA
292	CFH	Intron 10	TAGGTAGTCA-TATTTGGAAC
373	CFH	Intron 12	TGGACACATT-ATGATTGAGT
310	CFH	Exon 13	AGTTGGACCT-AATTCCGTTC
218	CFH	Intron 15	AGCTGAGTGA-CATGAGGTAG
382	CFH	Exon 18	GGAACCATTA-ATTCATCCAG
122	CFH	Exon 23	TCAATACATA-AATGCACCAA
164	CFHR3	Intron 1	AGGTAAGTTA-AAAGAGATCT
274	CFHR3	Intron 1	CATTTTCTTG-TGGAATTACA
391	CFHR3	Intron 3	CGGACGACAG-TCTCAGACTT
364	CFHR3	Intron 4	GGGTTATATG-AATTCCTACA
148	CFHR3	Intron 5	TTCCCCAACA-TCACAGCAGA
168	CFHR3	Exon 6	TCCCTTCCCG-ACACACTGCT
282	CFHR1	Intron 1	GGATAATTCA-ATTGAAATGG
346	CFHR1	Intron 3	AGAGTTTCAG-GTCCATGTGT
413	CFHR1	Intron 5	AATCTGTGAT-TATTTTGTTA
191	CFHR1	Exon 6	CCTGTTCTCA-AATAAAGCTT
454	CFHR1	Exon 6	TTTTCCAAGT-TTTAATATGG
406	CFHR2	Intron 1	AACTATGTCT-TGGAGTTTCG
265	CFHR2	Intron 2	AGATCATAAA-CACTTGATAA
226	CFHR2	Intron 3	AATACCTGTG-TGTGGTTTAT
184	CFHR2	Exon 4	ATGCTCCAGG-TTCATCAGTT
328	CFHR5	Exon 1	TGGGTATCCA-CTGTTGGGGG
427	CFHR5	Exon 2	TGAAGAAGAT-TATAACCCTT
232	CFHR5	Exon 3	CTTCAGGACT-AATACATCTG

***Table 7.7-1 Probes* used in MLPA assay**

7.8 Southern Blotting

7.8.1.1 Buffers

1. 20xSSC: 175.3 g sodium chloride and 88.2 g sodium citrate were dissolved in deionised water and made up to 1 L, pH 7.0 (with NaOH)
2. Denaturing buffer: 43.8 g sodium chloride added to 50 mL of 5 M sodium hydroxide diluted to 500 mL deionised water
3. Neutralisation buffer: 43.8 g sodium chloride added to 30.25 g Tris chloride dissolved in a total volume of 500 mL deionised water. HCl added to correct pH to 7.0
4. Wash buffer: 10 mL 20xSSC with 1g sodium dodecyl sulphate (Sigma) made up to 1 L with deionised water

7.8.1.2 Blot

10 µg genomic DNA was digested with the desired restriction enzyme in the appropriate buffer overnight. It was then subjected to electrophoresis at 80 V for 3 hours through an 0.8% agarose gel, photographed and exposed to ultraviolet radiation for 90 seconds. The gel was trimmed, a notch cut in the top left hand corner and left to soak in denaturing buffer for 30 minutes with gentle agitation. The gel was then rinsed in water and soaked in agitated neutralisation buffer for 30 minutes, changing the buffer after 15 minutes. A 'bridge' was fashioned using assorted plasticware and the gel placed on 3 Whatman 3M papers soaked in 20xSSC, with overhanging Whatman 3M paper dipping into a large reservoir of 20xSSC which was covered to prevent evaporation. A sheet of Saran wrap with a large rectangular hole very slightly smaller than the gel was placed over the gel and on top of the hole (and in contact with the gel and Saran wrap) was placed a piece of Hybond-N membrane (GE Healthcare, Little Chalfont, UK) which had been trimmed to the same size as the gel and soaked in deionised water for 5 minutes. On top of this was placed 3 Whatman 3M papers (trimmed to the same

size as the membrane) and then a 10 cm tower of paper towels under a glass plate and 0.5 kg mass. This contraption was left overnight to allow the DNA to transfer to the membrane.

The next morning, the set up was dismantled and notch cut in the top left hand corner of the membrane which was photographed (alongside the gel) to confirm transfer of DNA. The membrane was then rinsed with 2xSSC and allowed to air dry for < 1h on Whatman 3M paper. The membrane and Whatman 3M paper were then baked at 80°C for 1-2h to dryness and exposed to UV light for 3 minutes to denature the DNA.

7.8.1.3 Probe

DNA from a control individual was used as a template for PCR amplification of the desired probe sequence. The amplicon was run on a 2% agarose gel and gel purified. The purified amplicon was quantified and diluted to 25 ng in 45 µL in TE buffer (Sigma), denatured by heating to 95°C for 5 minutes before snap cooling on ice. It was then labelled with ^{32}P-dCTP (purchased from GE Healthcare) with random primers using the Rediprime II kit (GE Healthcare) according to the manufacturer's instructions. Unincorporated radioactive oligonucleotides were removed using NucAway spin columns (Applied Biosystems, Warrington, UK) according to the manufacturer's instructions.

7.8.1.4 Hybridization

The membrane was prehybridized in PerfectHyb buffer (Sigma) for 30 minutes at 60°C. 220 µL of purified, radiolabelled probe was added to 15 mL PerfectHyb buffer and the membrane incubated under this in a cylindrical chamber within a rotating hybridization oven at 68°C for 3 hours. The membrane was then washed in wash buffer at 60°C for 10 minutes. This was repeated a total of 3 times until radioactivity not readily detectable over the membrane. The membrane was then placed in an intensifying cartridge between acetate sheets and next to Kodak MXB autoradiography film (GRI, UK) and stored at -80°C for 3-5 days.

7.9 Plasmid generation and amplification

7.9.1 Antibiotic Stock (100 mg/mL)

1000x ampicillin and kanamycin stock solutions were prepared by dissolving 1 g of antibiotic powder in 10 mL of dH_2O. 500 µL aliquots of the stock solution were then prepared and stored at -20°C.

7.9.2 Luria-Bertani (LB) broth (Millers modification)

15.5 g of LB powder (Millers modification) was dissolved in 1 L of deionised H_2O and autoclaved. Antibiotics were added after sterilisation to produce selection media.

7.9.3 LB-agar plates

LB agar was prepared by placing 8.75 g of LB agar powder and 250 mL of deionised H_2O in a glass bottle. The mixture was then autoclaved to sterilise it and the LB agar was incubated in a water bath at 55°C before adding antibiotic at a final concentration of 100 µg/mL. Approximately 10 mL of LB agar was poured into each plastic bacterial culture dish and then allowed to set at room temperature, before being sealed with parafilm and stored at 4°C.

7.9.4 Transformation of E. coli

Chemically competent *Escherichia coli* (DH5α) cells were prepared by Tapan Bhattacharyya (Imperial College, London) and stored in aliquots at -80°C. Immediately prior to transformation DH5α were thawed on iced and 50 µL cells were transferred into a pre-chilled 15 mL polypropylene Falcon tube (Beckton Dickinson, Erembodegem, Belgium) for each transformation (and one control tube into which no plasmid DNA was to be added). 1-10 ng of plasmid DNA was added to the competent cells which were incubated on ice for 30 minutes. Next, the tubes were placed in a water bath at 42°C for 45 seconds to heat-shock the cells. The tubes were then placed immediately on ice for 2 minutes. 500 µL of LB broth (without antibiotics) at 42°C was added to each tube and the tubes were placed into a bacterial shaker

(Wolf laboratories, UK) for 1 hr at 37°C and 250 rpm. 100 µL of the LB mixture was subsequently spread onto an LB-Agar plate with antibiotic selection and incubated at 37°C overnight. Plates were inspected for colonies the following morning to identify whether transformation had occurred, and this was confirmed by checking that the plate inoculated with the control transformation (i.e. with no plasmid DNA) showed no colonies.

Single colonies were picked using a sterile pipette tip which was added to 3 mL of LB containing antibiotic (selection media) in 20 mL universal tubes (Bibby, UK). Cultures were grown overnight in a bacterial shaker (Wolf laboratories, UK) at 37°C, 150 rpm. DNA was then extracted using a QIAprep spin miniprep kit (Qiagen) according to the manufacturer's instructions. Briefly bacterial cells are pelleted, resuspended in buffer and lysed using alkaline lysis. The mixture was then neutralised and placed on a column containing a silica membrane to recover the DNA in high salt conditions. DNA was washed and eluted from the column to yield pure plasmid DNA which was verified either by restriction digestion or resequencing (Wolfson Institute of Biomedical Research, University College London).

7.9.5 Maxiprep of plasmid DNA

Following verification of the plasmid by restriction digestion or sequencing, a larger amount of DNA was prepared using a Qiaquick plasmid maxi prep kit (Qiagen). Colonies were picked from an LB-agar plate and cultured for 8 hrs in 2 mL of selection media. 250 µL of these cultures was used to seed a 250 mL culture of LB supplemented with antibiotic. The larger culture was then incubated overnight in a bacterial shaker (Wolf laboratories, UK) at 37°C, 150 rpm. Cells were pelleted and DNA extracted according to the manufacturer's instructions. This kit follows the same key steps described above.

7.9.6 Restriction digestion

DNA digestions with restriction enzymes were performed according to the manufacturer's instructions in the appropriate buffer (New England Biolabs, UK). The digested DNA was then analysed by agarose gel electrophoresis to confirm complete digestion.

7.9.7 DNA agarose gel electrophoresis

5x stock Tris Borate EDTA (TBE) buffer was made by mixing 54 g Tris(hydroxymethyl)aminomethane hydrochloride (Sigma, Poole, UK) with 27.5 g Boric acid (Sigma) and 100 mL 0.1M EDTA (Sigma) and the mixture made up to 1 L with dH_2O.

Agarose gels were prepared by diluting agarose powder in 1xTBE buffer to the desired concentration (0.8% w/v to separate fragments >1kbp, 1-2% for fragments between 200bp and 1kbp and 3% to separate products of similar size, such as those resulting from ARMS-PCR). The mixture was heated until boiling in a microwave oven and ethidium bromide was added to the molten agarose immediately prior to casting to yield a final concentration of 0.5 µg/mL. Samples were mixed with 6x loading dye (New England Biolabs) and loaded on the agarose gel alongside the appropriate DNA size ladders (100 bp ladder, 1 kb ladder; New England Biolabs). A typical electrophoresis run was performed under 1xTBE at a constant 100 V. DNA bands were visualised and photographed under ultraviolet light (Syngene, Cambridge).

7.9.8 Alkaline phosphatase treatment

1 unit of Calf intestinal alkaline phosphatase (Invitrogen, UK) was added to 50 µL restriction digestion and incubated at 37^{o}C for 5 minutes. The alkaline phosphatase was inactivated by heating to 65^{o}C for 15 minutes.

7.9.9 Gel purification

DNA bands of the desired size were resected from agarose gel following electrophoresis using a scalpel under ultraviolet light. The DNA was extracted using a QIAquick gel extraction kit (QIAGEN, Germany) according to the manufacturer's instructions.

7.9.10 Ligation

Ligation was performed using T4 DNA ligase (Promega) using gel-purified linearised, phosphatased vector and gel-purified digested insert which were quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). Molecular equivalents (MEs) were calculated by dividing the measured concentration (in ng/µL) by the size of the DNA molecule (in nucleotides) and the desired ratios made up in the reaction mix according to the manufacturer's instructions. The mixture was incubated at 16^{o}C for 16 hours.

7.9.11 Generation of mutant HIF2α *alleles*

pcDNA3.1 constructs containing the human HIF2α coding sequence with and without a C-terminal MYC epitope tag were kindly supplied by P Maxwell (UCL) and these were used for functional studies.

7.9.11.1 Mutagenesis

Variants encoding HIF2αGly537→Arg and Gly537→Trp were generated using Quikchange (Stratagene, La Jolla CA) according to the manufacturer's instructions. Briefly, 33-mer primers were designed incorporating the desired (i.e. mutant) nucleotide at position 16-18 (***Table 7.9-1***).

Gene	Name	Sequence (5' → 3')
HIF2A	537 Arg F	CCTATATCCCCATGGAC**A**GGGAAGACTTCCAGC
HIF2A	537 Arg R	GCTGGAAGTCTTCCC**T**GTCCATGGGGATATAGG
HIF2A	537 Trp F	CCTATATCCCCATGGAC**T**GGGAAGACTTCCAGC
HIF2A	537 Trp R	GCTGGAAGTCTTCCC**A**GTCCATGGGGATATAGG
CFHR5	C5ex2-3 F	ACTCCTCCCATATGCAGCTTCACTAGAACACTTTGTG
CFHR5	C5ex2-3 R	GGCTGAGCATCTACATTGGCTTCTAAAATTGGAACATGAC

Table 7.9-1 Primers used for the generation of mutant constructs
Single nucleotide substitutions are shown in red

10 ng of pcDNA3.1HIF2α plasmid DNA was used as a template for a PCR reaction which used *PfuUltra* HF DNA polymerase and the following program ***Table 7.9-2***:

Step	Temperature	Time
1	95°C	30 s
2	95°C	30 s
3	55°C	1 min
4	68°C	1 min/kb plasmid length

Table 7.9-2 PCR program for Quikchange mutagenesis
Steps 2 - 4 were repeated 16 times

The reaction mixture was then digested with *DpnI* at 37°C for 1 hour to destroy the (methylated) template DNA. The nicked (mutant) amplicon plasmids were then used to transform chemically competent cells.

7.10 Cell culture

All tissue culture reagents described were supplied by Gibco (Invitrogen, UK) unless otherwise stated.

HepG2 hepatocellular carcinoma cells were a kind gift from Bhathiya Wijeyekoon (Oxford University). The basal media used was RPMI 1640 supplemented with 10% foetal calf serum (FCS), 2 mM L-Glutamine, 50 IU/mL penicillin and 50 µg/mL streptomycin. Stock cell cultures were maintained in 100 mm tissue culture dishes (TPP, UK) in a GalaxyR CO_2 incubator (Wolf Laboratories, UK) with a humidified atmosphere of 5% CO_2 at 37^{o}C. To passage cells, monolayers were washed with RPMI 1640 with no supplements and then dissociated with trypsin-EDTA solution. Complete medium was added and cells were transferred to a 50 mL Falcon tube. Cells were then spun at 1000 rpm for 5 minutes in a centrifuge (Hettich Zentrifugen, Tuttlingen) to pellet the cells. After spinning the supernatant was removed and the pellet resuspended into an appropriate volume of complete medium before re-plating into fresh tissue culture dishes.

7.10.1 Preparation of frozen cell stocks

Cells from a confluent 100 mm tissue culture dish were trypsinised as described above. After pelleting, cells were resuspended in 3 mL of FCS. 3 mL of 2x freezing medium (20% dimethyl sulphoxide (DMSO), 80% FCS) was then added to the cells in a dropwise manner. 2 mL aliquots of cells were prepared in cryovials. The cryovials were placed in a -80^{o}C freezer overnight before being transferred to liquid nitrogen for long-term storage. Cells were defrosted by hand warming until the cryovial contained a slurry. This was emptied into a 50 mL Falcon tube containing complete medium. The cells were then pelleted and resuspended in 1 mL complete medium before adding them to tissue culture plates.

7.10.2 Transfection of cultured cells for HRE-reporter assay

HepG2 cells in 6-well plates were co-transfected using Lipofectamine 2000 (Invitrogen, Paisley UK) according to the manufacturer's instructions. All experiments were performed in triplicate on at least 2 separate occasions. In brief, 5.5 µL Lipofectamine 2000 was added to 244.5 µL Optimem and incubated for 5 minutes. This was then added to a mixture containing: the desired amount of the expression plasmid; 2 µg of pGL3PGK6TKpLUC reporter construct (which incorporates a multimerised Hypoxia Response Element (HRE) linked to thymidine kinase-driven Firefly *Luciferase* gene(Vaux, Wood et al. 2001)); 150 ng pRL-SV40 containing *Renilla Luciferase* (Promega, Southampton UK) made up to a volume of 250 µL with Optimem. After 20 minutes, this 500 µL mixture was added to cell monolayers in 6-well plates under 2.5 mL of RPMI supplemented with 10% FCS and 2 mM L-Glutamine but no antibiotics.

Hypoxic incubation was performed using a GalaxyR tissue culture incubator (Wolf Laboratories, UK). CO_2 was fixed at 5% and hypoxia was maintained with supplemental N_2.

48 hours after transfection, cells were lysed and luminescence measured using Stop and Glo (Promega) according to the manufacturer's instructions. Luminescence was measured using a TD-20/20 Luminometer (Turner Designs, Sunnyvale CA).

Appendix C: Consent Form and Information Sheet

Participant Consent Form

Title of project:

Genetic Investigation of Kidney Disease

The participant should complete the whole of this sheet him or herself.

(please tick each statement if it applies to you)

I have read the Information Sheet for Patients and Healthy Relatives ☐

I have been given the opportunity to ask questions and discuss this study ☐

I have received enough information about the study ☐

The study has been explained to me by:
Prof/Dr/Mr/Mrs/Ms_____

I understand that I am free to withdraw from the study at any time, without having to give a reason for withdrawing and without affecting my future medical care ☐

I agree to my general practitioner and/or nephrologist being informed of my participation in this study ☐

I agree for my genetic material to be extracted, analysed and stored for this study ☐

Please tick this box if you would like to be informed when a result is available ☐

Signed.....Date.....

(NAME IN BLOCK CAPITALS).....

Investigator's signature.....Date: .....

(NAME IN BLOCK CAPITALS).....

Information Sheet for Patients and Healthy Relatives

You will be given a copy of this Information Sheet

1. Study title
Genetic Investigation of Kidney Disease

2. Invitation
You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

3. What is the purpose of the study?
We want to understand why kidney disease particularly affects you and/or your family.

4. Why have I been chosen?
You have been chosen for this study because your family appears to have a particularly high incidence of kidney disease and this is likely to have a genetic cause.

5. Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

6. What will happen to me if I take part?
Agreeing to take part in the study will mean that we will need to take a small extra sample of blood (about a tablespoonful). Alternatively, a small sample of saliva can be used instead, although we prefer to use blood. If you agree to take part it means that we will look at your DNA for changes that might explain why you and/or members of your family have developed kidney disease.
In most cases, we expect that we will not be able to find an explanation at the current time for the increased risk of kidney disease. This is because we do not yet understand all of the genes that are involved, and because other, non-genetic factors may also contribute. If we did find out anything that we think explains the increased incidence of kidney disease in your family, we will discuss this with you, and explain what we think it means.

7. **What are the possible benefits of taking part?**
We may be able to understand why you or members of your family have developed (or might be at increased risk of) kidney disease.

8. **What if something goes wrong?**
If you are harmed by taking part in this research project, you will be compensated through the Imperial College School of Medicine's "No Fault" Compensation Scheme.

9. **Will my taking part in this study be kept confidential?**
All information which is collected about you during the course of the research will be kept strictly confidential. Any information about you which leaves the hospital will have your name and address removed so that you cannot be recognized from it. Please indicate on the consent form if you do **not** want your general practitioner and/or nephrologist to be informed of your participation in this study.

10. **What will happen to the results of the research study?**
Provided the research goes as planned the results will be published in a scientific journal. If you would like us to tell you when we publish something please indicate this by ticking the box on the form.

11. **What will happen to the samples after the research has been done?**
DNA samples extracted from your blood will be stored in the Renal Laboratory so that other researchers can use them for further research into kidney disease. Researchers from outside the renal department will not have access to your name or personal details. If you do not want DNA from your blood to be stored in this way, please inform the researcher.

12. **Who is organising and funding the research?**
Professor Maxwell is organising the research. The resources required are met by the Medical Research Council.

13. **Who has reviewed the study?**
The study has been reviewed by the HAMMERSMITH, QUEEN CHARLOTTE'S & CHELSEA AND ACTON HOSPITALS Research Ethics Committee.

14. **Contact for Further Information**
If you would like any further information at any time please contact Dr Daniel Gale, UCL Division of Medicine, 5 University Street London, WC1E 6JF; Email: d.gale@ucl.ac.uk

Thank you for taking the time to read this information sheet. If you feel able to take part in the study we are very grateful to you.

Appendix D: Papers arising from this work

1. Daniel P. Gale, Sarah K. Harten, Cecil D. L. Reid, Edward G. D. Tuddenham and Patrick H. Maxwell. Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating HIF2{alpha} mutation. **Blood** 2008;112: 919-921

2. Daniel P Gale, Elena Goicoechea de Jorge, H Terence Cook, Rubén Martinez-Barricarte, Andreas Hadjisavvas, Adam G McLean, Charles D Pusey, Alkis Pierides, Kyriacos Kyriacou, Yiannis Athanasiou, Konstantinos Voskarides, Constantinos Deltas, Andrew Palmer, Véronique Frémeaux-Bacchi, Santiago Rodriguez de Cordoba, Patrick H Maxwell, Matthew C Pickering. Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis. **The Lancet** 2010;376(9743): 794 – 801.

blood
2008 112: 919-921
doi:10.1182/blood-2008-04-153718

Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating HIF2{alpha} mutation

Daniel P. Gale, Sarah K. Harten, Cecil D. L. Reid, Edward G. D. Tuddenham and Patrick H. Maxwell

Updated information and services can be found at:
http://bloodjournal.hematologylibrary.org/cgi/content/full/112/3/919

Information about reproducing this article in parts or in its entirety may be found online at:
http://bloodjournal.hematologylibrary.org/misc/rights.dtl#repub_requests

Information about ordering reprints may be found online at:
http://bloodjournal.hematologylibrary.org/misc/rights.dtl#reprints

Information about subscriptions and ASH membership may be found online at:
http://bloodjournal.hematologylibrary.org/subscriptions/index.dtl

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published semimonthly by the American Society of Hematology, 1900 M St, NW, Suite 200, Washington DC 20036.
Copyright 2007 by The American Society of Hematology; all rights reserved.

Figure 1. Frequencies of *BCR-ABL* in B-precursor ALL according to age groups. Fifteen (0.6%) patients showed both types of transcripts (M-bcr and m-bcr) and were not included in this diagram.

investigated for *BCR-ABL* transcripts with 2 different polymerase chain reaction (PCR) methods as previously described in detail.[1,2] Forty-six (1.8%) cases showed an ambiguous result and were not considered for further analysis. Of the remaining 2498 cases, 904 (36.2%) were *BCR-ABL* (599 24% minor breakpoint region [m-bcr] and 282 11.3% major breakpoint region [M-bcr], 15 0.6% both, M-bcr and m-bcr, and 8 atypical transcripts), and 1594 were *BCR-ABL* . Atypical transcripts were not systematically detected before 2000[1] and thus had to be excluded from further analysis. We grouped patients into age cohorts at 10-year intervals according to their age at diagnosis, each comprising between 277 and 481 patients and found a remarkable increase of *BCR-ABL* frequency in adolescents and young adults (Figure 1). It increased from 12.7% in adolescents (15-24 years) to 30.6% and 43.7% in patients aged 25 to 34 and 35 to 44 years, respectively. In patients older than 44 years, the *BCR-ABL* frequency showed no further increment and ranged between 42% and 44%. The increase of *BCR-ABL* frequency was paralleled by a relative increase of M-bcr transcripts. These transcripts accounted for 16.4% of all *BCR-ABL*-positive cases in adolescents (15-24 years). Their relative frequency increased to 22.5% in 25- to 34-year-olds and to 36.8% in 35- to 44-year-olds and remained between 33% and 36.2% from then on.

The reason for this age dependency is not obvious. The relative frequencies of immunologic subtypes (78.2% common, 19.9% pre-B, 1.9% pro-B) of *BCR-ABL* patients did not differ significantly across the age groups. Moreover, the frequency of *BCR-ABL* was also not significantly different in woman compared with men. Our study excluded lymphatic blast crises in patients with known chronic myeloid leukemia (CML). Previous work has indicated that M-bcr- and m-bcr-positive ALL may arise from different sets of hematopoietic progenitor cells,[3] but this does not explain the age dependency. A number of genetic markers in ALL show a marked age dependency (reviewed in Armstrong and Look[4]), eg, *TEL-AML1*, *TLX1*, and *TLX3*, *MLL* aberrations (especially in infant ALL). Our data provide additional information on the biology of *BCR-ABL*-positive ALL and substantiate evidence of age-dependent variation in the genetic background of ALL.

Thomas Burmeister, Stefan Schwartz, Claus R. Bartram, Nicola Gokbuget, Dieter Hoelzer, and Eckhard Thiel ¨

This work was supported in part by the Deutsche Krebshilfe (German Cancer Aid), Bonn, Germany.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Dr Thomas Burmeister, Charite Universitatsmedizin Berlin, Campus Benjamin Franklin, Medizinische Klinik III, Hindenburgdamm 30, 12200 Berlin, Germany; e-mail: thomas.burmeister@charite.de.

References

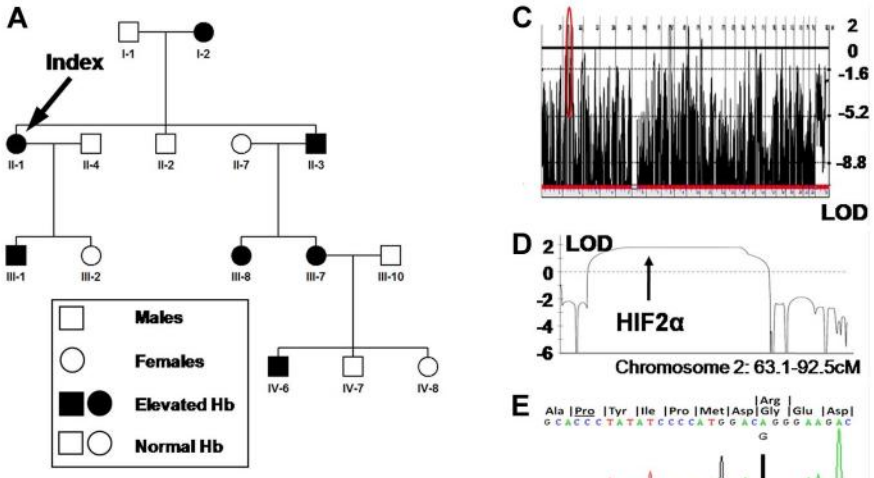
1. Burmeister T, Schwartz S, Taubald A, et al. Atypical BCR-ABL mRNA transcripts in adult ALL. Haematologica. 2007;92:1699-1702.
2. Gleissner B, Gokbuget N, Bartram CR, et al. Leading prognostic relevance of the BCR-ABL translocation in adult acute B-lineage lymphoblastic leukemia: a prospective study of the German Multicenter Trial Group and confirmed polymerase chain reaction analysis. Blood. 2002;99:1536-1543.
3. Castor A, Nilsson L, Astrand-Grundstrom I, et al. Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia. Nat Med. 2005; 11:630-637.
4. Armstrong SA, Look AT. Molecular genetics of acute lymphoblastic leukemia. J Clin Oncol. 2005;23:6306-6315.

To the editor:

Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating *HIF2* mutation

Erythropoietin production is regulated by the transcription factor hypoxia-inducible factor (HIF). Erythrocytosis with raised erythropoietin levels due to dysregulated HIF activity is recognized as a consequence of a hypomorphic HIF E3 ubiquitin ligase (*VHL*) allele (Chuvash polycythemia), inactivating mutations of the HIF hydroxylase *PHD2* and, very recently, activating mutations of *HIF2* .[1-3] We report that erythrocytosis in a large kindred originally reported in 1979[4] is due to an activating *HIF2* mutation. Two affected individuals developed severe pulmonary hypertension, a hitherto unrecognized consequence of mutations in the pathway.

This study was approved by the Hammersmith and Queen Charlotte's and Chelsea Hospitals local research ethics committee. Informed consent was obtained in accordance with the Declaration of Helsinki. At presentation, affected family members (Figure 1A) exhibited erythrocytosis (hemoglobin up to 228 g/L) with elevated total red cell volume and reduced plasma volume (Figure 1B). Other hematological parameters, serum biochemistry, arterial



B

Parameter	Normal Range (adult)	II-1	II-3	I-2	III-7
Hemoglobin (g/L)	*115-175*	202*	189*	228	197
EPO (milli-IU/mL)	*9.1-23*	>2000*	1980*	–	26.2
Total red cell/plasma volume (mL/kg)	*22-32/ 40-50*	33/28	56.6/–	34/22.4	–
Pulmonary Hypertension		Present	Present	–	–

Figure 1. Data and findings for selected family members. (A) Family tree. (B) Hematological parameters in selected family members. I-2 was diagnosed with "heart failure" at age 62; echocardiography was not available. *Not contemporaneous measurements. - indicates data not available. (C) Genome-wide linkage to autosomal dominant erythrocytosis. (D) Parametric LOD score at region circled in red in panel C. (E) Sequencing trace of showing G/A heterozygosity at base 2097, predicting a change from glycine to arginine at residue 537, 6 codons from critical hydroxyl acceptor proline (underlined). (F) HRE-Firefly:*Renilla* luciferase activity following cotransfection of a plasmid containing the coding sequence of HIF2 with or without mutations shown. Results shown are means plus or minus SD of 3 independent experiments, each performed in triplicate.

oxygenation, and oxygen-hemoglobin dissociation were normal. Serum erythropoietin was elevated, rising to more than 2000 milli-IU/mL. II-1 and II-3 both had documented pulmo- nary arterial (PA) hypertension in their sixth decade without any evidence of thromboembolism. A genome-wide single nucleo- tide polymorphism screen was compatible with linkage to HIF2 (Logarithm or odds ratio [LOD] 1.81, Figure 1C,D) and resequencing disclosed heterozygosity for a G 3 A substitution at position 2097 (Figure 1E), predicting a glycine-to-arginine change at residue 537 of the protein (HIF2 Arg537), present in all affected and no unaffected individuals. Expression of HIF2 Arg537 in HepG2 (a hepatoma cell line) cells revealed increased activation of a hypoxia response element-luciferase reporter construct compared with wild-type (Figure 1F). HIF2 Arg537 was more active than HIF2 Trp537, consistent with a more severe phenotype in this family.

These findings confirm that an activating HIF2 mutation dysregulates erythropoietin production in humans. That HIF2 plays a central role in regulating erythropoietin is supported by several observations. First, in hepatoma and neuroblastoma cell lines, erythropoietin mRNA was suppressed by siRNA-mediated silencing of HIF2 , but not HIF1 .[5] Second, in *Vhl*$^{R/R}$ mice homozygous for a mutation equivalent to that underlying Chuvash polycythemia in humans, there was erythrocytosis in normoxia associated with stabilization of HIF2 , but not HIF1 .[6] Third, reduced oxygen delivery to the kidney in rats activates HIF2 , but not HIF1 in the fibroblasts which produce erythropoietin.[7]

This report provides the first evidence that genetic HIF activation could cause severe pulmonary hypertension in later life. This is plausible; although clinically important pulmonary hypertension is not evident in Chuvash polycythemia, there is a detectable increase in resting PA pressure and an exaggerated hypoxic pulmonary vasoconstriction response.[8,9] There is also experimental support for a critical role of HIF2 in the pulmonary vasculature since *hif2* mice are protected from right ventricular hypertrophy in hypoxia.[10]

These findings have several implications. First, there should be a high index of suspicion for pulmonary hypertension in other kindreds with activation of the HIF pathway. Second, inhibitors of PHD enzymes, which are in late stage clinical trials for treatment of anemia, may cause pulmonary hypertension. Third, it raises the possibility that polymorphic variation in HIF2 contributes to the marked differential susceptibility to erythrocytosis, reduced plasma volume, and pulmonary hypertension in humans at high altitude.

Daniel P. Gale, Sarah K. Harten, Cecil D. L. Reid, Edward G. D. Tuddenham, and Patrick H. Maxwell

The authors are grateful to the family and the family's physicians, and thank Dr C. Mein for performing the SNP genotyping and Mr T. Bhattacharyya for technical assistance.

This work was supported by a UK Medical Research Council Fellowship (D.P.G.) and a British Heart Foundation programme grant (P.H.M.).

Contribution: D.P.G. and P.H.M. designed research, analyzed data, and wrote the paper. D.P.G. and S.K.H. carried out research. E.G.D.T. and C.D.L.R. contributed, looked after the patients, and read and approved the paper.

Conflict-of-interest disclosure: P.H.M. is a scientific founder and director of ReOx, Ltd, which aims to develop PHD inhibitors. The remaining authors declare no competing financial interests.

Correspondence: Patrick Maxwell, Division of Medicine, The Rayne Institute, 5 University Street, London WC1E 6JF, United Kingdom; e-mail: p.maxwell@ucl.ac.uk.

References

1. Ang SO, Chen H, Hirota K, et al. Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. Nat Genet. 2002;32:614-621.
2. Percy MJ, Zhao Q, Flores A, et al. A family with erythrocytosis establishes a role for prolyl hydroxylase domain protein 2 in oxygen homeostasis. Proc Natl Acad Sci U S A. 2006;103:654-659.
3. Percy MJ, Furlow PW, Lucas GS, et al. A gain-of-function mutation in the HIF2A gene in familial erythrocytosis. N Engl J Med. 2008;358:162-168.
4. Howarth C, Chanarin I, Janowska-Wieczorek A, et al. Familial erythrocytosis. Scand J Haematol. 1979;23:217-222.
5. Warnecke C, Zaborowska Z, Kurreck J, et al. Differentiating the functional role of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha (EPAS-1) by the use of RNA interference: erythropoietin is a HIF-2alpha target gene in Hep3B and Kelly cells. FASEB J. 2004;18:1462-1464.
6. Hickey MM, Lam JC, Bezman NA, et al. von Hippel-Lindau mutation in mice recapitulates Chuvash polycythemia via hypoxia-inducible factor-2alpha signaling and splenic erythropoiesis. J Clin Invest. 2007;117:3879-3889.
7. Wiesener MS, Jurgensen JS, Rosenberger C, et al. Widespread hypoxia-inducible expression of HIF-2alpha in distinct cell populations of different organs. FASEB J. 2003;17:271-273.
8. Smith TG, Brooks JT, Balanos GM, et al. Mutation of von Hippel-Lindau tumour suppressor and human cardiopulmonary physiology. PLoS Med. 2006;3:e290.
9. Bushuev VI, Miasnikova GY, Sergueeva AI, et al. Endothelin-1, vascular endothelial growth factor and systolic pulmonary artery pressure in patients with Chuvash polycythemia. Haematologica. 2006;91:744-749.
10. Brusselmans K, Compernolle V, Tjwa M, et al. Heterozygous deficiency of hypoxia-inducible factor-2alpha protects mice against pulmonary hypertension and right ventricular dysfunction during prolonged hypoxia. J Clin Invest. 2003;111:1519-1527.

Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis

Daniel P Gale*, Elena Goicoechea de Jorge*, H Terence Cook, Rubén Martinez-Barricarte, Andreas Hadjisavvas, Adam G McLean, Charles D Pusey, Attis Pierides, Kyriacos Kyriacou, Yiannis Athanasiou, Konstantinos Voskarides, Constantinos Deltas, Andrew Palmer, Véronique Frémeaux-Bacchi, Santiago Rodriguez de Cordoba, Patrick H Maxwell†, and Matthew C Pickering†

Lancet 2010; 376: 794–801
Published Online August 26, 2010
DOI:10.1016/S0140-6736(10)60670-8

See Comment page 748

*Both authors contributed equally to the report
†Joint senior authors
Division of Medicine, University College, London, UK (D P Gale MRCP, Prof P H Maxwell FRCP); Centre for Complement and Inflammation Research (E Goicoechea de Jorge PhD, Prof M C Pickering PhD, Prof H T Cook FRCPath) and Imperial College Kidney and Transplant Institute (D P Gale, A G McLean FRCP, Prof C D Pusey FRCP, A Palmer FRCP), Imperial College, London, UK; Centro de Investigaciones Biológicas (CSIC), Centro de Investigación Biomédica en Enfermedades Raras and Instituto Reina Sofía de Investigaciones Nefrológicas, Madrid, Spain (R Martinez-Barricarte PhD, Prof S Rodriguez de Cordoba PhD); Department of Electron Microscopy and Pathology, Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus (A Hadjisavvas PhD, K Kyriacou PhD); Department of Nephrology, Hippocrateon Hospital, Nicosia, Cyprus (A Pierides FRCP); Department of Nephrology, Nicosia General Hospital, Nicosia, Cyprus (Y Athanasiou MD); Department of Biological Science, University of Cyprus, Nicosia, Cyprus (K Voskarides PhD, Prof C Deltas PhD); Assistance Publique-Hopitaux de Paris, Service d'Immunologie Biologique, Hôpital Européen Georges Pompidou, Paris, France (V Frémeaux-Bacchi PhD); and INSERM UMRS 872, Cordeliers Research Centre, Paris, France (V Frémeaux-Bacchi)

Summary

Background Complement is a key component of the innate immune system, and variation in genes that regulate its activation is associated with renal and other disease. We aimed to establish the genetic basis for a familial disorder of complement regulation associated with persistent microscopic haematuria, recurrent macroscopic haematuria, glomerulonephritis, and progressive renal failure.

Methods We sought patients from the West London Renal and Transplant Centre (London, UK) with unusual renal disease and affected family members as a method of identification of new genetic causes of kidney disease. Two families of Cypriot origin were identified in which renal disease was consistent with autosomal dominant transmission and renal biopsy of at least one individual showed C3 glomerulonephritis. A mutation was identified via a genome-wide linkage study and candidate gene analysis. A PCR-based diagnostic test was then developed and used to screen for the mutation in population-based samples and in individuals and families with renal disease.

Findings Occurrence of familial renal disease cosegregated with the same mutation in the complement factor H-related protein 5 gene (*CFHR5*). In a cohort of 84 Cypriots with unexplained renal disease, four had mutation in *CFHR5*. Overall, we identified 26 individuals with the mutation and evidence of renal disease from 11 ostensibly unrelated kindreds, including the original two families. A mutant CFHR5 protein present in patient serum had reduced affinity for surface-bound complement. We term this renal disease CFHR5 nephropathy.

Interpretation CFHR5 nephropathy accounts for a substantial burden of renal disease in patients of Cypriot origin and can be diagnosed with a specific molecular test. The high risk of progressive renal disease in carriers of the *CFHR5* mutation implies that isolated microscopic haematuria or recurrent macroscopic haematuria should not be regarded as a benign finding in individuals of Cypriot descent.

Funding UK Medical Research Council and Wellcome Trust.

Introduction

Kidney disease is an important cause of morbidity and mortality worldwide. In many cases, renal injury results from damage caused by the immune system, either in response to microbial infection or because of inappropriate activation of defence mechanisms. The mechanisms that protect the kidney from immunological attack in healthy individuals—and that fail in disease—are not well understood.

The complement system is a key component of host defence, and variation in the genes that regulate complement activation is associated with disease, including age-related macular degeneration,[1,2] atypical haemolytic uraemic syndrome,[3,4] and glomerulonephritis.[2,5–7] The kidney is especially susceptible to the effects of complement activation, and glomerulonephritis (a leading cause of kidney failure worldwide) is generally characterised by presence of complement within the glomerulus. Typically, complement is accompanied by immunoglobulins, which activate it via the classical pathway. However, complement deposition can occur without immunoglobulin via the complement alternative pathway. This deposition occurs in dense-deposit disease, which is caused by genetic or acquired defects in complement regulation.[5]

Isolated glomerular C3 deposition and inflammation can also arise in the absence of dense-deposit disease. This heterogeneous entity has been termed C3 glomerulonephritis and is often associated with the histological appearance of membranoproliferative glomerulonephritis.[7] Our aim was to investigate an inherited renal disease, which we show is endemic in Cyprus and is characterised by microscopic and synpharyngitic macroscopic haematuria, renal failure, and C3 glomerulonephritis.

Methods

Patients

To detect high penetrance genes leading to kidney disease, we identified multiply affected kindreds of patients from the West London Renal and Transplant Centre (London, UK), prioritising those with an unusual renal condition, syndromic features, or early onset of disease. Family 1 in this report lived in London, UK, and reported ancestry from the Troodos mountains of Cyprus. The index patient from family 2 was referred to us from

Cyprus with C3 glomerulonephritis and, because he also came from the Troodos region and C3 glomerulonephritis is very rare, we postulated that he might have the same genetic condition as individuals from family 1. Individuals from both families were tested for evidence of renal disease and underwent genetic analysis, leading to identification of a shared mutation.

To establish the frequency of this genetic mutation, we searched for carriers in two cohorts. We examined DNA for 102 unrelated individuals in the UK 1958 birth cohort[8] and 1015 control participants in the MASTOS study in Cyprus.[9]

We sought additional individuals in Cyprus by screening for the presence of the mutation in a cohort of 84 Cypriot patients with advanced or end-stage chronic renal disease, either of unknown cause or because of presumed or incompletely characterised glomerulonephritis. A further two families from Cyprus (family 3 and family 4) were tested for the mutation because they had familial renal disease in which other conditions had been sought and excluded.[10] In these families, microscopic and synpharyngitic macroscopic haematuria segregated as an autosomal dominant trait and direct exon sequencing excluded recognised mutations of *COL4A3* and *COL4A4*.[10]

We sought additional individuals in London, UK, by reviewing case records from the West London Renal and Transplant Centre and the Royal Free Renal Unit for potentially affected individuals, looking for histological features consistent with CFHR5 nephropathy. This search identified four further individuals who were shown to have the mutation. One patient had a brother with microscopic haematuria, who was also found to have the mutation (family 5).

Finally, we analysed DNA from 36 individuals with a renal biopsy diagnosis of C3GN (defined as glomerular C3 deposition in the absence of immunoglobulin staining for IgG, IgM, and IgA, and without intramembranous glomerular basement membrane dense deposits) and no recognised family history; 34 individuals were from France (including 19 reported previously[7]) and two were from the UK. Additional single nucleotide polymorphisms (SNPs) were genotyped in three members of five families (CytoSNP 12 panel, Illumina, CA, USA) to define the maximum possible extent of the shared haplotype spanning the locus. Estimation of the probable number of generations since the common affected ancestor of the families was done as previously described.[11] Formal retinal examination was done by an ophthalmologist from Moorfields Eye Hospital, London, UK for two individuals (IV-5 and V-4) from family 1.

The study was approved by the local research ethics committees and participants provided written consent.

Genetic analysis

Individuals from families 1 and 2 were tested for evidence of renal disease, and DNA was extracted from blood or saliva (Oragene, DNA Genotek, Kanata, ON, Canada). Genotypes and haplotypes of 6008 SNPs (Linkage IV panel, Illumina, CA, USA) from these two families were analysed with EasyLINKAGE,[12] PEDCHECK,[13] GENEHUNTER version 2.1,[14] and HAPLOPAINTER.[15] Bidirectional sequencing of the exons of candidate genes was done after PCR amplification (primers available from the authors on request).

CFHR5 internal duplication was assessed by multiplex ligation-dependent probe amplification (MLPA), which was done with unamplified genomic DNA with the P236 A1 ARMD mix 1 (MRC-Holland, Amsterdam, Netherlands). Southern blotting was done with genomic DNA digested with *Eco*R1 (New England Biolabs, MA, USA). Membranes were probed with a ^{32}P-labelled 371 base pair sequence containing exon 2 of *CFHR5*. PCR amplification of the *CFHR5* duplication insertion point used the primers 5′-TGGAAGCCTGTGGTATAAATGA-3′ and 5′-TCCGGCACATCCTTCTCTAT-3′. Screening PCR to amplify both *CFHR5* alleles in a single reaction used the primers 5′-GATTCCATTTGTCAAATATTG-3′, 5′-TCTTCTCCAAAACTATCTAATGTCAA-3′, and 5′-TTTGAATGCTGTTTTAGCTCG-3′.

Correspondence to:
Prof Patrick Maxwell, Division of Medicine, The Rayne Institute, 5 University Street, London WC1E 6JF, UK
p.maxwell@ucl.ac.uk

For the multiplex ligation-dependent probe amplification procedure see http://www.mrc-holland.com

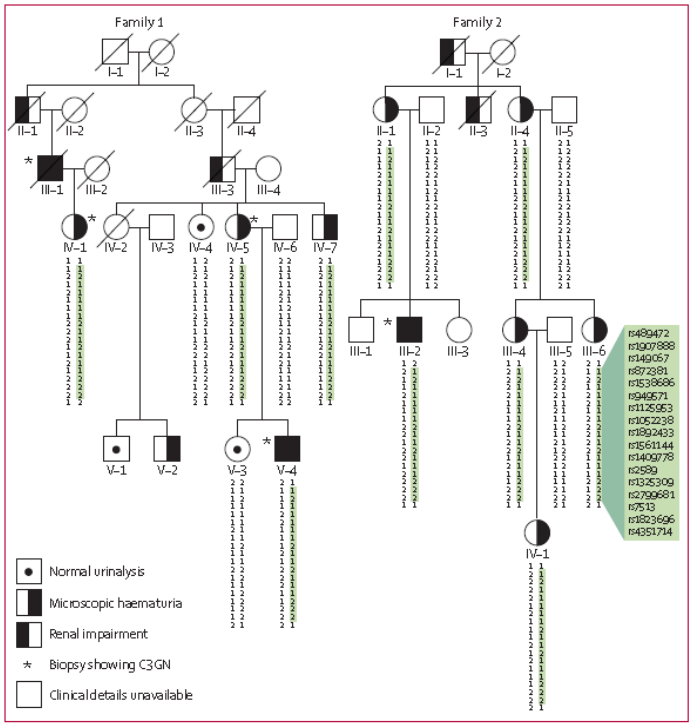


Figure 1: **Pedigrees of two ostensibly unrelated families from the Troodos mountain region, Cyprus, showing haplotypes at region of maximum linkage on chromosome 1**
Green boxes are haplotypes (HapMap coordinates 192344247 to 201027281).

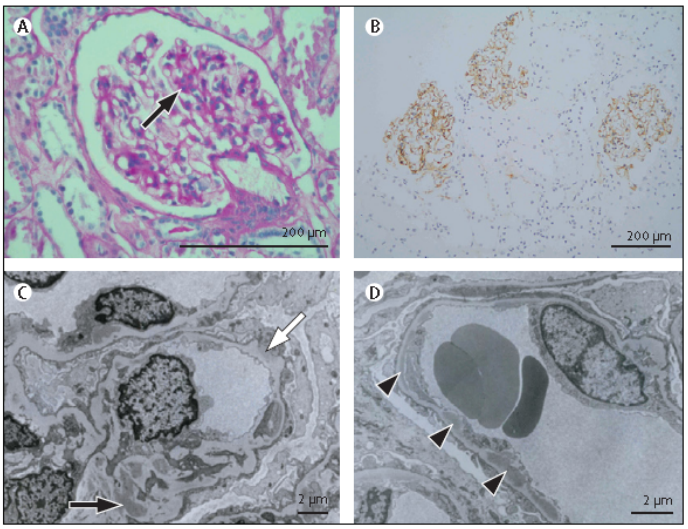


Figure 2: **Renal biopsy sample of patient V-4 from family 1**
(A) Periodic acid-Schiff stain of glomerulus demonstrating mesangial hypercellularity (arrow). (B) Immunoperoxidase stain for C3 showing granular staining on the capillary wall. (C and D) Electron micrographs showing prominent mesangial electron-dense deposits (black arrow), subendothelial deposits (arrowheads), and occasional subepithelial deposits (white arrow).

Serum CFHR5 detection and functional analysis

We used western blotting to detect CFHR5 in serum and recombinant CFHR5 in supernatants using a rabbit polyclonal anti-human-CFHR5 antibody[16] (a gift from J McRae, Immunology Research Centre, Melbourne, Australia). Functional analysis of CFHR5 protein binding to heparin and lysed chicken erythrocytes was done as previously described.[16,17] Briefly, patient serum was incubated with chicken erythrocytes, which spontaneously activated the alternative pathway resulting in cell lysis and binding of CFHR5 to the disrupted membranes. The relative amounts of unbound CFHR5 protein (supernatant) and bound CFHR5 protein (erythrocyte membrane pellet) were established by western blotting.

Role of the funding source

The funding bodies of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. PHM and MCP had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Renal disease segregated as an autosomal dominant trait in family 1 and family 2 (figure 1). Disease was characterised by persistent microscopic haematuria and episodes of syntharyngitic macroscopic haematuria (within 1–2 days of an upper respiratory tract infection). C3 glomerulonephritis was identified in renal biopsy specimens in affected male and female patients. Glomerular inflammation was variable, but subendothelial and mesangial deposits that were reactive with antibody to C3, but not to C1q, C4, or immunoglobulins, were always present in these patients. Subepithelial deposits were occasionally noted. Figure 2 shows a representative biopsy sample.

In family 1, a genome-wide SNP-based analysis that used an autosomal dominant model and scored all patients with haematuria as affected established linkage to an 18 cM (centimorgan) region of chromosome 1q31–32 (logarithm of the odds [LOD] score 2·22). With the addition of family 2, the combined LOD score was 3·40 (figure 3). A haplotype of 17 SNPs spanning 8·74 cM within the linked region was shared by all affected members of both families, which is consistent with inheritance of an allele at this locus from a common ancestor (figure 1). This haplotype (HapMap coordinates 192344247 to 201027281) included complement factor H (*CFH*) and CFH-related (*CFHR*) 1–5 genes. Sequencing of *CFH* and two of the CFH-related genes (*CFHR1* and *CFHR5*) in patients V-4 and IV-5 (both family 1) did not reveal any mutations.

MLPA analysis detected a heterozygous deletion of *CFHR1* and *CFHR3* in three members of family 2 (figure 4), which did not segregate with renal disease and is a polymorphism present in 14·2% of healthy controls.[18] MLPA analysis also showed heterozygosity for a previously unreported duplication of exons 2 and 3 of *CFHR5* in affected individuals from family 1 (IV-5 and V-4) and family 2 (II-1 and III-2), but not in the unaffected individual in family 2 (II-2) (figure 4). The *CFHR5* internal duplication was confirmed, and its size established, by Southern hybridisation of genomic DNA with *CFHR5* exon 2. The boundary of the duplication was identified by resequencing a PCR product (figure 5). Figure 5 shows an additional 6·3 kbp fragment on Southern blot analysis in affected individuals from family 1 (IV-5 and V-4) and family 2 (II-1 and III-2). This band was not present in two unrelated control individuals or in an unaffected individual from family 2 (II-2), in whom only the expected 7·9 kbp band was seen. With a reverse primer positioned in *CFHR5* exon 2 and a forward primer in intron 3, there was no amplification from wild-type DNA (as expected). The duplication of exons 2 and 3 in DNA with the mutation led to a product of 4·8 kbp with these primers. Sequencing of this product confirmed the position and size of the duplication. A diagnostic PCR test was then designed with three primers which gave a 298 bp product from the wild-type allele and an additional 222 bp product from the allele with the *CFHR5* internal duplication. The 222 bp product was seen in affected but not unaffected members of families 1 and 2 (figure 6).

With this diagnostic PCR, we did not detect a *CFHR5* internal duplication in 102 randomly selected participants from the UK 1958 birth cohort.[8] The *CFHR5* internal duplication was only detected in one individual of 1015 control participants in the MASTOS study (for

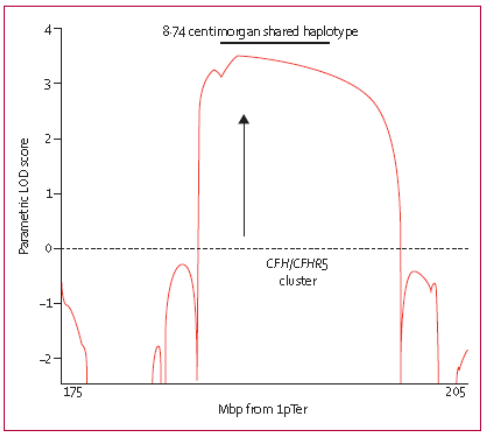


Figure 3: **Genome-wide single-nucleotide polymorphism-based analysis**
LOD=logarithm of the odds.

whom no clinical or personal details were available).[9] The *CFHR5* internal duplication is therefore a rare allele in the Cypriot population.

We then screened a cohort of 84 patients in Cyprus with advanced or end-stage renal failure. This screening identified one female and three male patients with the mutation, none of whom reported ancestry in the Troodos mountains. Families 3 and 4 were identified by testing Cypriot families with unexplained haematuria, and the *CFHR5* duplication was identified in all seven individuals affected with microscopic or sympharyngitic macroscopic haematuria, but not in five unaffected relatives. A renal biopsy had been taken from one individual from these families and showed C3 glomerulonephritis. Neither of these families could trace their ancestry to the Troodos mountains region. The review of case records from two renal units in London followed by DNA testing identified three men and one woman of Cypriot descent with CFHR5 nephropathy. In one of these cases a brother with microscopic haematuria was also tested and found to have the *CFHR5* mutation (family 5). Although not ostensibly related to family 1 or family 2, paternal ancestry was from the same valley in the Troodos Mountains.

The table summarises available clinical data for all the individuals we showed had the *CFHR5* mutation or who were obligate carriers. In all affected men, there was progressive renal impairment with end-stage renal disease between the ages of 40 and 69 years. Renal impairment was less common in affected women than in affected men (p=0·0016; table).

These findings suggest that the newly identified *CFHR5* duplication accounts for a substantial proportion of renal disease in Cyprus and is not confined to the Troodos mountain region. Further genotyping of SNPs showed a haplotype extending from 1·4 cM upstream to 4·7 cM downstream of the mutation that cosegregated with the *CFHR5* internal duplication in the five families tested, which is consistent with inheritance of the mutation by all affected individuals from an original founder. By contrast, none of the 36 individuals with biopsies showing C3 glomerulonephritis from the UK and France had the *CFHR5* mutation.

CFHR5 is a 65 kDa plasma protein consisting of nine short consensus repeat (SCR) domains (repeating units that are characteristic of CFH and CFHR proteins).[10] Duplication of exon 2 (encoding SCR1) and exon 3 (encoding SCR2), predicted a novel CFHR5 protein with duplicated SCR domains 1 and 2 (figure 7). Western

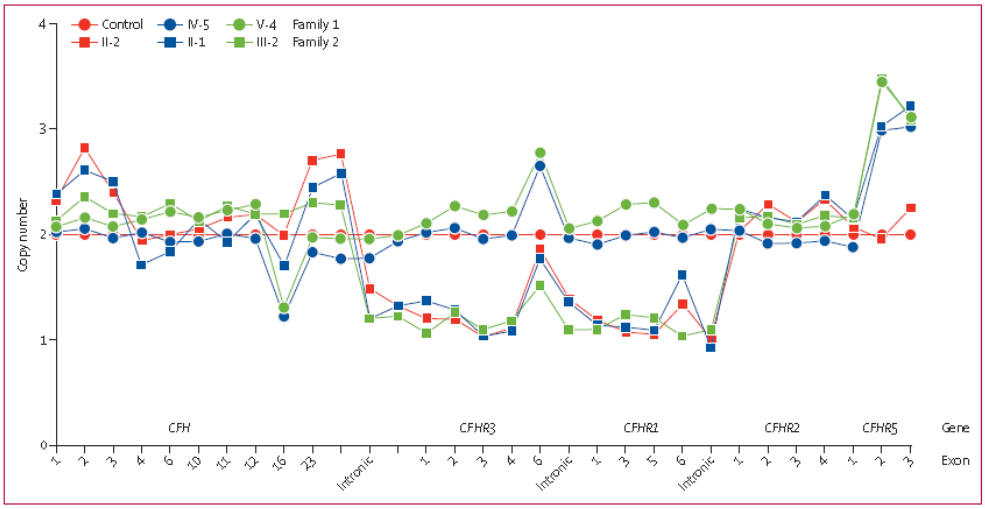


Figure 4: **Multiplex ligation-dependent probe amplification analysis of *CFH* and *CFHR* genes using genomic DNA from five individuals from families 1 and 2**

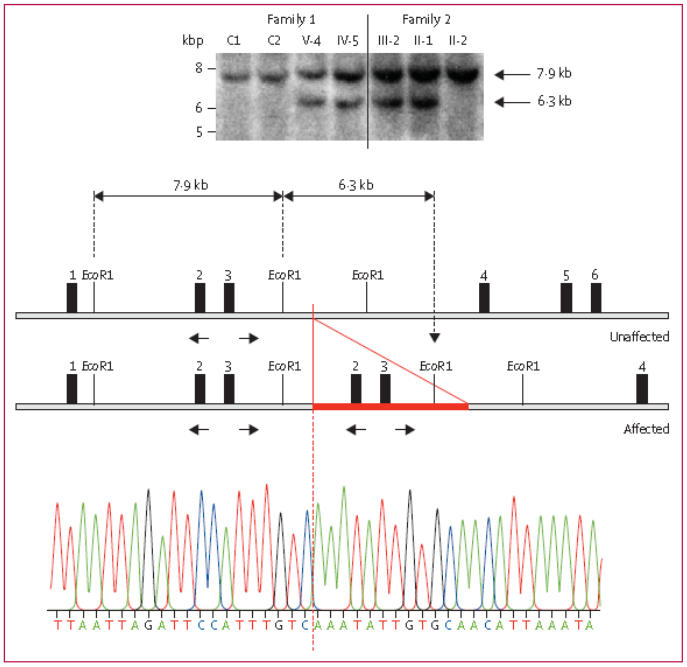


Figure 5: **Southern blot of *EcoR1*-digested genomic DNA from two unrelated controls (C1 and C2) and five individuals from families 1 and 2, probed with exon 2 of *CFHR5***
Arrows show orientation of PCR primers. Numbered black rectangles are CFHR5 exons. Double-headed arrows are EcoR1 restriction fragments. Red triangle shows the duplicated region.

See Online for webappendix

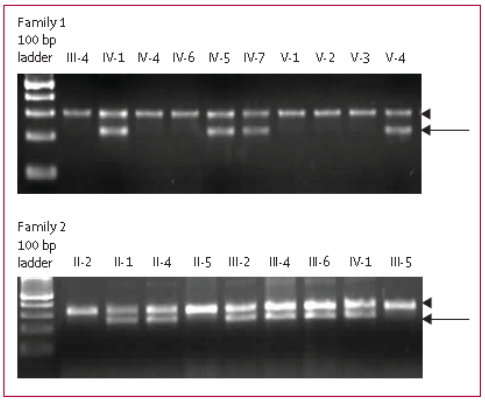


Figure 6: **Diagnostic PCR test for *CFHR5*$^{12123\text{-}9}$**
Arrowhead shows one 298 bp product in the wild-type allele and arrows shows the 222 bp product in the allele associated with CFHR5 internal duplication.

blotting of serum samples from affected individuals detected the normal CFHR5$^{123\text{-}9}$ (superscript numbers denote SCRs) protein, and a slow migrating protein that was consistent with the predicted molecular weight of CFHR5$^{12123\text{-}9}$ (figure 7).

Although the initial aminoacid sequence of the wild-type SCR1 is Glu-Gly-Thr-Leu-Cys-Asp (with the first glutamic acid encoded by exon 1), the duplicated SCR1 is encoded only by exon 2 and therefore is predicted not to have this glutamic acid. Splicing of exon 3 upstream of the exon 2 sequence results in a change in the initial codon of the duplicated SCR1 so that the second aminoacid in the sequence is arginine rather than glycine. Hence, the initial aminoacid sequence of the duplicated SCR1 is predicted to be Arg-Thr-Leu-Cys-Asp. The rest of the duplicated SCR1 aminoacid sequence and entire duplicated SCR2 aminoacid sequence is identical to wild-type CFHR5.

The physiological role of CFHR5 is not known; the plasma concentration is estimated to be 3–6 μg/mL,[19] which is approximately 1% of that of the well characterised complement regulator CFH.[20] CFHR5 has been shown to co-localise with complement deposits in diseased human kidneys[21] and binds to surface-bound activated C3 (termed C3b).[16] We showed that mutant CFHR5$^{12123\text{-}9}$ protein bound less effectively than did wild-type protein to two sources of surface-bound complement; complement-lysed erythrocyte membranes (figure 8) and glomerular-bound mouse complement (webappendix p 1).

Previous in-vitro investigations have reported that CFHR5 shares some of the complement regulatory activities of CFH; CFHR5 can be a cofactor for the proteolytic inactivation of complement C3b by the plasma enzyme complement factor I, although its activity is weak by comparison with that of CFH.[16] We confirmed that both the wild-type CFHR5 and mutant CFHR5$^{12123\text{-}9}$ proteins have complement factor I cofactor activity; in both cases the activity is substantially weaker than that reported for CFH. The mutant CFHR5$^{12123\text{-}9}$ protein did not have reduced activity compared with wild-type in this assay; perhaps surprisingly, complement factor I cofactor activity was increased (webappendix p 2).

Discussion

We provide evidence for an inherited renal disease, endemic in Cyprus, that is characterised by microscopic and synpharyngitic macroscopic haematuria, renal failure, and C3 glomerulonephritis, and show that affected individuals have an internal duplication within the gene for complement factor H-related protein 5 (*CFHR5*). We term this disease CFHR5 nephropathy.

Isolated microscopic haematuria is a common presentation that in the absence of urinary tract abnormalities, proteinuria, or renal impairment is usually believed to be benign, and with present guidelines is not investigated by renal biopsy.[22,23] However, this study shows that isolated microscopic haematuria can be the presenting feature of progressive renal disease, with implications for patients and their families. Furthermore, analysis of data suggests that familial isolated microscopic haematuria attributable to heterozygous mutations in *COL4A3* and *COL4A4* is also associated with progressive

chronic kidney disease.[19] Taken together, these findings underscore the importance of any history of renal disease in the family and the value of renal biopsy and genetic investigations in this setting.

CFHR5 nephropathy has several noteworthy clinical features. First, the risk of progressive renal impairment is more common in men than in women. Second, since Berger's original description,[24] repeated episodes of synpharyngitic haematuria have been regarded as almost diagnostic of IgA nephropathy; CFHR5 nephropathy should now be thought of as a differential diagnosis. Third, although retinopathy is a well recognised feature of dense-deposit disease, with affected individuals developing ocular drusen at a young age, clinically significant visual impairment was not a feature of CFHR5 nephropathy. This finding is consistent with a study[25] that reported no association between CFHR5 polymorphisms and risk of age-related macular degeneration. However, because formal ophthalmological assessment was only done for two affected individuals in this study, we cannot exclude the presence of subclinical ocular disease in CFHR5 nephropathy.

An important pathological consideration was that most of the individuals who were biopsied initially had a histological diagnosis of membranoproliferative glomerulonephritis type 1. Review led to recognition that these abnormalities were C3 glomerulonephritis, which is an entity described in a French report[7] of 19 patients that did not identify families with more than one affected individual. In that series,[7] men were not more severely affected than were women, synpharyngitic macroscopic haematuria was not reported, and microscopic haematuria was absent in seven of 19 (37%) patients at diagnosis. By contrast, CFHR5 nephropathy is an inherited disease, characterised by synpharyngitic macroscopic haematuria (noted in seven [58%] of the affected individuals tested; see table) and microscopic haematuria (all affected individuals tested). Thus, CFHR5 nephropathy constitutes a distinct clinical entity that is associated with the histological category C3 glomerulonephritis. Abnormalities in other complement genes were identified in a subset of the patients in the French report,[7] suggesting that different mechanisms of complement dysregulation can lead to much the same histological appearances.

The *CFHR5*^{121239} mutation is a copy number variation (ie, a duplication or deletion of a segment of DNA), which is not detectable with standard exon-based sequencing. Copy number variations arise through non-allelic recombination events that are more common in regions of complex genomic architecture such as the *CFH* and *CFHR1–5* gene cluster.

The extent of the shared haplotype suggests that affected members of the families inherited the mutation from a common ancestor about 16 generations ago,[11] and our identification of additional cases and affected families in the Cypriot population implies the existence of a substantial number of individuals with *CFHR5*^{121239}. The high penetrance of haematuria (all 22 mutation carriers tested; see table), the wide geographical distribution of ancestry within Cyprus, and the presence of affected individuals in the UK suggests that this disease will account for a substantial proportion of renal disease affecting inhabitants of the island and their descendants worldwide. Population-based studies need to be done to quantify this proportion; we examined only a small cohort of Cypriots with advanced or end-stage renal disease, either of unknown cause or attributed to presumed or incompletely characterised glomerulonephritis. In the US renal data system, about 25% of patients with advanced renal disease would fit into this category.

Although CFHR5 nephropathy seems to be a common cause of renal disease in the Cypriot population, we did not detect this specific *CFHR5* mutation in 36 cases of sporadic C3 glomerulonephritis from France and the UK. Whether this geographical disparity is the effect of ascertainment bias, positive selection for the allele within Cyprus (perhaps because of the presence of an endemic infectious disease), or genetic drift within the island's population is not known. So far, mutations in other complement regulatory genes and the presence of C3 nephritic factor have been reported in a few cases of C3 glomerulonephritis;[7] other genetic abnormalities of *CFHR5* might exist.

The mutation we identified in *CFHR5* provides a robust genetic marker for a novel hereditary nephritis, and screening for the mutation is a reliable clinical test.

	Male	Female	Total
Individuals with CFHR5 mutation	14	12	26
With microscopic haematuria	11/11	11/11	22/22
Obligate carriers with impaired renal function	3/3	0/0	3/3
Confirmed affecteds and obligate carriers with impaired renal function*	13/16†	2/12	15/28
Affecteds with synpharyngitic macroscopic haematuria	5/8	2/4	7/12
Biopsy findings in affected individuals			
Mesangial matrix expansion	6/6	2/3	8/9
Increased glomerular cellularity	6/6	2/3	8/9
Segmental capillary wall thickening	6/6	2/3	8/9
Glomerular staining for C3	6/6	3/3	9/9
Glomerular staining for C1q, IgA, IgG, or IgM	0/6	0/3	0/9
Subendothelial electron-dense deposits	6/6	3/3	9/9
Mesangial electron-dense deposits	6/6	3/3	9/9
Scanty subepithelial electron-dense deposits	4/6	3/3	7/9
CFHR5^{121239} detected in serum	7/7	4/4	11/11
Normal serum C3, C4, CFH, and CFI	9/9	5/5	14/14
Negative for C3NeF	2/2	2/2	4/4
Normal retinal appearance‡	1/1	1/1	2/2
Self-reported visual impairment	0/8	0/7	0/15
Additional ungenotyped first-degree relatives with impaired renal function	5	0	5

Data are number/number of confirmed affecteds and obligate carriers of *CFHR5* duplication for whom data were available. CFHR5=complement factor H-related protein 5. CFH=complement factor H. CFI=complement factor I. C3NeF=C3 nephritic factor. *p=0·0016 men versus women, Fisher's exact test. †One patient's DNA and urine was analysed, but no data on renal function were available. ‡From fundoscopic examination by ophthalmologist.

Table: **Clinical and pathological findings in CFHR5 nephropathy**

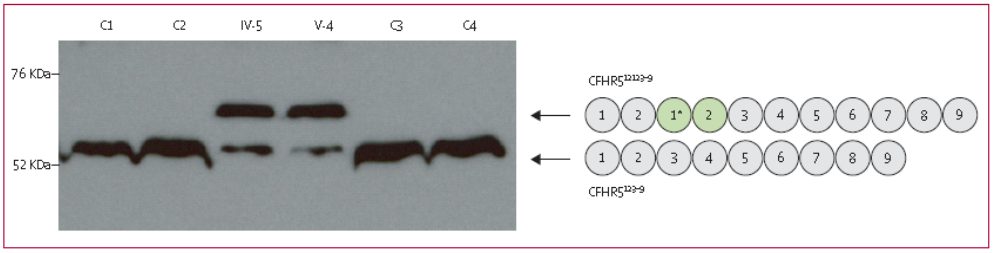


Figure 7: **Western blot of serum with a polyclonal anti-CFHR5 antibody for detection of CFHR5**
In affected patients there is a slower migrating band consistent with the predicted molecular weight of CFHR5$^{12123\text{-}9}$. Duplicated short consensus repeat domains are shown in green. *Initial amino acid sequence in the duplicate SCR1 differs from the original SCR1, as described in the text.

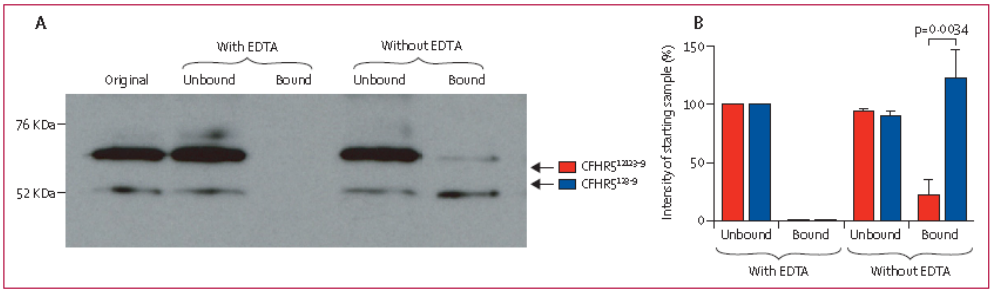


Figure 8: **Functional analysis of CFHR5 proteins**
(A) Western blot of bound and unbound CFHR5 to complement-lysed erythrocyte membranes. (B) Quantitative data from three experiments analysing binding to complement-lysed erythrocyte membranes.

However, our understanding of how the mutation causes the disease is incomplete and we do not exclude the possibility that another mutation within the 6·1 cM-shared haplotype could exist. We propose that CFHR5 is important in complement processing within the kidney, and that the *CFHR5*$^{12123\text{-}9}$ mutation impairs the ability of CFHR5 to achieve this. Our model is consistent with the previous finding[21] that wild-type CFHR5 co-localises with complement within the kidney in renal disease.

Our in-vitro functional studies show that the mutant CFHR5$^{12123\text{-}9}$ protein binds to surface-bound complement (on erythrocytes and mouse glomeruli) substantially less well than does wild-type CFHR5. Western blot analysis of serum from affected individuals showed greater intensity for mutant CFHR5$^{12123\text{-}9}$ protein than for wild-type CFHR5 (figure 7); an in-vivo finding that would be consistent with a defect in recruitment of the mutant protein from the circulation to interact with surface-bound complement. However, *CFHR5*$^{12123\text{-}9}$ is not a straightforward loss of functional allele, because it produces a circulating protein that shows enhanced complement factor I cofactor activity in vitro compared with wild-type CFHR5.

In laboratory models of C3 dysregulation, excessive production of inactivated C3b (iC3b) by complement factor I is important for the initiation of renal injury;[5] complement factor I cofactor activity of CFHR5$^{12123\text{-}9}$ could lead to an increase in iC3b in the glomerulus. More studies will be necessary to elucidate fully the biological role of CFHR5 and the pathophysiology of C3 glomerulonephritis in CFHR5 nephropathy. The central role of abnormal complement deposition in this disease suggests that inhibition of the terminal complement pathway (for instance with the humanised anticomplement C5 monoclonal antibody eculizumab, which is effective for treatment of paroxysmal nocturnal haemoglobinuria[26]) might be of therapeutic benefit in CFHR5 nephropathy and clinical studies are needed to address this issue.

Contributors
DPG, EGdJ, PHM, and MCP designed the study. DPG, EGdJ, HTC, RM-B, AH, KV, YA, APi, and CD did the investigation. Additional assistance was provided by APa, CDP, AGM, KK, VF-B, and SRdC. DPG, EGdJ, HTC, PHM, and MCP interpreted the data, and DPG, PHM, and MCP wrote the report with the help of all authors.

Conflicts of interest
Imperial College, London, UK and University College, London, UK have a patent pending on CFHR5 in renal disease filed as a consequence of this work. We declare that we have no conflicts of interest.

Acknowledgments
DPG is supported by the UK Medical Research Council and EGdJ and MCP are supported by the Wellcome Trust. Additional support was provided by the UK National Institute for Health Research Biomedical Research Centre Funding Scheme and the Cyprus Research Promotion Foundation. No payment was received for the writing of this article. We thank the patients and their families. DPG is supported by a UK Medical Research Council Clinical Research Training Fellowship. MCP is a Wellcome Trust Senior Fellow in Clinical Science (WT082291MA) and EGdJ is funded by this fellowship. CD is supported by the Cyprus Research Promotion Foundation through grants ΕΝΙΣΧ/0505/02 and ΕΝΙΣΧ/0308/08. Additional support was obtained from the UK National

Institute for Health Research Biomedical Research Centre Funding Scheme. PHM is supported by the EU large scale collaborative project Metoxia, the St Peter's Trust, and a Senior Investigator Award from the UK National Institute for Health Research.

References

1 Hageman GS, Anderson DH, Johnson LV, et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci USA* 2005; **102:** 7227–32.

2 de Cordoba SR, Goicoechea de Jorge E. Translational mini-review series on complement factor H: genetics and disease associations of human complement factor H. *Clin Exp Immunol* 2008; **151:** 1–13.

3 Goicoechea de Jorge E, Harris CL, Esparza-Gordillo J, et al. Gain-of-function mutations in complement factor B are associated with atypical hemolytic uremic syndrome. *Proc Natl Acad Sci USA* 2007; **104:** 240–45.

4 Fremeaux-Bacchi V, Miller EC, Liszewski MK, et al. Mutations in complement C3 predispose to development of atypical hemolytic uremic syndrome. *Blood* 2008; **112:** 4948–52.

5 Pickering MC, Cook HT. Translational mini-review series on complement factor H: renal diseases associated with complement factor H: novel insights from humans and animals. *Clin Exp Immunol* 2008; **151:** 210–30.

6 Smith RJ, Alexander J, Barlow PN, et al. New approaches to the treatment of dense deposit disease. *J Am Soc Nephrol* 2007; **18:** 2447–56.

7 Servais A, Fremeaux-Bacchi V, Lequintrec M, et al. Primary glomerulonephritis with isolated C3 deposits: a new entity which shares common genetic risk factors with haemolytic uraemic syndrome. *J Med Genet* 2007; **44:** 193–99.

8 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447:** 661–78.

9 Loizidou MA, Michael T, Neuhausen SL, et al. Genetic polymorphisms in the DNA repair genes *XRCC1*, *XRCC2* and *XRCC3* and risk of breast cancer in Cyprus. *Breast Cancer Res Treat* 2008; **112:** 575–79.

10 Voskarides K, Damianou L, Neocleous V, et al. *COL4A3/COL4A4* mutations producing focal segmental glomerulosclerosis and renal failure in thin basement membrane nephropathy. *J Am Soc Nephrol* 2007; **18:** 3004–16.

11 Genin E, Tullio-Pelet A, Begeot F, Lyonnet S, Abel L. Estimating the age of rare disease mutations: the example of Triple-A syndrome. *J Med Genet* 2004; **41:** 445–49.

12 Hoffmann K, Lindner TH. easyLINKAGE-Plus—automated linkage analyses using large-scale SNP data. *Bioinformatics* 2005; **21:** 3565–67.

13 O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63:** 259–66.

14 Markianos K, Daly MJ, Kruglyak L. Efficient multipoint linkage analysis through reduction of inheritance space. *Am J Hum Genet* 2001; **68:** 963–77.

15 Thiele H, Nurnberg P. HaploPainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics* 2005; **21:** 1730–32.

16 McRae JL, Cowan PJ, Power DA, et al. Human factor H-related protein 5 (FHR-5). A new complement-associated protein. *J Biol Chem* 2001; **276:** 6747–54.

17 Pickering MC, de Jorge EG, Martinez-Barricarte R, et al. Spontaneous hemolytic uremic syndrome triggered by complement factor H lacking surface recognition domains. *J Exp Med* 2007; **204:** 1249–56.

18 Dragon-Durey MA, Blanc C, Marliot F, et al. The high frequency of complement factor H related *CFHR1* gene deletion is restricted to specific subgroups of patients with atypical haemolytic uraemic syndrome. *J Med Genet* 2009; **46:** 447–50.

19 McRae JL, Duthy TG, Griggs KM, et al. Human factor H-related protein 5 has cofactor activity, inhibits C3 convertase activity, binds heparin and C-reactive protein, and associates with lipoprotein. *J Immunol* 2005; **174:** 6250–56.

20 Whaley K, Ruddy S. Modulation of the alternative complement pathways by beta 1 H globulin. *J Exp Med* 1976; **144:** 1147–63.

21 Murphy B, Georgiou T, Machet D, Hill P, McRae J. Factor H-related protein-5: a novel component of human glomerular immune deposits. *Am J Kidney Dis* 2002; **39:** 24–27.

22 Cohen RA, Brown RS. Clinical practice. Microscopic hematuria. *N Engl J Med* 2003; **348:** 2330–38.

23 National Collaborating Centre for Chronic Conditions. Chronic kidney disease: national clinical guideline for early identification and management in adults in primary and secondary care. London: Royal College of Physicians, 2008: 147–48.

24 Berger J, Hinglais N. Les dépôts intercapillaires d'IgA-IgG. *J Urol Nephrol* 1968; **74:** 694-95.

25 Narendra U, Pauer GJ, Hagstrom SA. Genetic analysis of complement factor H related 5, CFHR5, in patients with age-related macular degeneration. *Mol Vis* 2009; **15:** 731–36.

26 Hillmen P, Young NS, Schubert J, et al. The complement inhibitor eculizumab in paroxysmal nocturnal hemoglobinuria. *N Engl J Med* 2006; **355:** 1233–43.

References

Abarrategui-Garrido, C., R. Martinez-Barricarte, et al. (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**(19): 4261-4271.

Abecasis, G. R., S. S. Cherny, et al. (2002). "Merlin--rapid analysis of dense genetic maps using sparse gene flow trees." Nat Genet **30**(1): 97-101.

Abrera-Abeleda, M. A., C. Nishimura, et al. (2006). "Variations in the complement regulatory genes factor H (CFH) and factor H related 5 (CFHR5) are associated with membranoproliferative glomerulonephritis type II (dense deposit disease)." J Med Genet **43**(7): 582-589.

Ahearn, J. M. and D. T. Fearon (1989). "Structure and function of the complement receptors, CR1 (CD35) and CR2 (CD21)." Adv Immunol **46**: 183-219.

Aitman, T. J., R. Dong, et al. (2006). "Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans." Nature **439**(7078): 851-855.

Al-Sheikh, M., K. Moradkhani, et al. (2008). "Disturbance in the HIF-1alpha pathway associated with erythrocytosis: further evidences brought by frameshift and nonsense mutations in the prolyl hydroxylase domain protein 2 (PHD2) gene." Blood Cells Mol Dis **40**(2): 160-165.

Alam, H., E. T. Maizels, et al. (2004). "Follicle-stimulating hormone activation of hypoxia-inducible factor-1 by the phosphatidylinositol 3-kinase/AKT/Ras homolog enriched in brain (Rheb)/mammalian target of rapamycin (mTOR) pathway is necessary for induction of select protein markers of follicular differentiation." J Biol Chem **279**(19): 19431-19440.

Albertini, A. M., M. Hofer, et al. (1982). "On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions." Cell **29**(2): 319-328.

Alhopuro, P., D. Phichith, et al. (2008). "Unregulated smooth-muscle myosin in human intestinal neoplasia." Proc Natl Acad Sci U S A **105**(14): 5513-5518.

Amadei, N., G. V. Baracho, et al. (2001). "Inherited complete factor I deficiency associated with systemic lupus erythematosus, higher susceptibility to infection and low levels of factor H." Scand J Immunol **53**(6): 615-621.

Amara, U., D. Rittirsch, et al. (2008). "Interaction between the coagulation and complement system." Adv Exp Med Biol **632**: 71-79.

Ang, S. O., H. Chen, et al. (2002). "Endemic polycythemia in Russia: mutation in the VHL gene." Blood Cells Mol Dis **28**(1): 57-62.

Ang, S. O., H. Chen, et al. (2002). "Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia." Nat Genet **32**(4): 614-621.

Ansell, D., T. Feest, et al. (2004). "UK Renal Registry 7th Annual Report."

Arias-Stella, J., H. Kruger, et al. (1973). "Pathology of chronic mountain sickness." Thorax **28**(6): 701-708.

Asokan, R., J. Hua, et al. (2006). "Characterization of human complement receptor type 2 (CR2/CD21) as a receptor for IFN-alpha: a potential role in systemic lupus erythematosus." J Immunol **177**(1): 383-394.

Baracho, G. V., V. Nudelman, et al. (2003). "Molecular characterization of homozygous hereditary factor I deficiency." Clin Exp Immunol **131**(2): 280-286.

Barbour, S. J., L. Er, et al. (2010). "Differences in progression of CKD and mortality amongst Caucasian, Oriental Asian and South Asian CKD patients." Nephrol Dial Transplant.

Barrett, J. C., B. Fry, et al. (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics **21**(2): 263-265.

Barry, R. E. and W. Krek (2004). "The von Hippel-Lindau tumour suppressor: a multi-faceted inhibitor of tumourigenesis." Trends Mol Med **10**(9): 466-472.

Beall, C. M. and A. B. Reichsman (1984). "Hemoglobin levels in a Himalayan high altitude population." Am J Phys Anthropol **63**(3): 301-306.

Beck, I., R. Weinmann, et al. (1993). "Characterization of hypoxia-responsive enhancer in the human erythropoietin gene shows presence of hypoxia-inducible 120-Kd nuclear DNA-binding protein in erythropoietin-producing and nonproducing cells." Blood **82**(3): 704-711.

Ben Nasr, A., J. Haithcoat, et al. (2006). "Critical role for serum opsonins and complement receptors CR3 (CD11b/CD18) and CR4 (CD11c/CD18) in phagocytosis of Francisella tularensis by human dendritic cells (DC): uptake of Francisella leads to activation of immature DC and intracellular survival of the bacteria." J Leukoc Biol **80**(4): 774-786.

Bennett, A., S. R. Sain, et al. (2008). "Evidence that parent-of-origin affects birth-weight reductions at high altitude." Am J Hum Biol **20**(5): 592-597.

Bert, P. (1878). La pression barometrique: recherches de physiologie experimentale. Paris, Masson.

Bilzer, M., F. Roggel, et al. (2006). "Role of Kupffer cells in host defense and liver disease." Liver Int **26**(10): 1175-1186.

Boon, C. J., N. C. van de Kar, et al. (2009). "The spectrum of phenotypes caused by variants in the CFH gene." Mol Immunol **46**(8-9): 1573-1594.

Bordet, J. (1895). "Les leukocytes et les proprietes actives du serum chez les vaccines." Ann. Inst. Pasteur (Paris) **9**: 462-506.

Boyer, O., L. H. Noel, et al. (2008). "Complement factor H deficiency and posttransplantation glomerulonephritis with isolated C3 deposits." Am J Kidney Dis **51**(4): 671-677.

Boyle, W. S. and G. E. Smith (1970). "Charge Coupled Semiconductor Devices." Bell Sys. Tech. J. **49**(4): 587-593.

Bushuev, V. I., G. Y. Miasnikova, et al. (2006). "Endothelin-1, vascular endothelial growth factor and systolic pulmonary artery pressure in patients with Chuvash polycythemia." Haematologica **91**(6): 744-749.

Byrne, C., D. Ford, et al. (2009). "UK Renal Registry 12th Annual Report (December 2009): Chapter 3: UK ESRD incident rates in 2008: national and centre-specific analyses." Nephron Clin Pract **115 Suppl 1**: c9-39.

Caprioli, J., M. Noris, et al. (2006). "Genetics of HUS: the impact of MCP, CFH, and IF mutations on clinical presentation, response to treatment, and outcome." Blood **108**(4): 1267-1279.

Carnot, P. and C. Deflandre (1906). "Sur l'activite haemopoietique du serum." Comptes rendus hebdomadaires des séances de l'Académie des sciences **143**: 432-435.

Cedzynski, M., J. Szemraj, et al. (2004). "Mannan-binding lectin insufficiency in children with recurrent infections of the respiratory system." Clin Exp Immunol **136**(2): 304-311.

Chambers, J. C., W. Zhang, et al. (2009). "Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels." Nat Genet **41**(11): 1170-1172.

Chang, J. T. (1999). "Recent Common Ancestors of All Present-Day Individuals." Advanced Applied Probability **31**: 1002-1026.

Clemenza, L. and D. E. Isenman (2000). "Structure-guided identification of C3d residues essential for its binding to complement receptor 2 (CD21)." J Immunol **165**(7): 3839-3848.

Cockburn, I. A., M. J. Mackinnon, et al. (2004). "A human complement receptor 1 polymorphism that reduces Plasmodium falciparum rosetting confers protection against severe malaria." Proc Natl Acad Sci U S A **101**(1): 272-277.

Cohen, R. A. and R. S. Brown (2003). "Clinical practice. Microscopic hematuria." N Engl J Med **348**(23): 2330-2338.

Colville, D., R. Guymer, et al. (2003). "Visual impairment caused by retinal abnormalities in mesangiocapillary (membranoproliferative) glomerulonephritis type II ("dense deposit disease")." Am J Kidney Dis **42**(2): E2-5.

Conrad, D. H., J. R. Carlo, et al. (1978). "Interaction of beta1H globulin with cell-bound C3b: quantitative analysis of binding and influence of alternative pathway components on binding." J Exp Med **147**(6): 1792-1805.

Constantinou, K. (1998). "Anopheles (malaria) eradication in Cyprus." Parassitologia **40**(1-2): 131-135.

Cotes, P. M. and D. R. Bangham (1961). "Bio-assay of erythropoietin in mice made polycythaemic by exposure to air at a reduced pressure." Nature **191**: 1065-1067.

Cottingham, R. W., Jr., R. M. Idury, et al. (1993). "Faster sequential genetic linkage computations." Am J Hum Genet **53**(1): 252-263.

Cringle, S. J. and D. Y. Yu (2004). "Intraretinal oxygenation and oxygen consumption in the rabbit during systemic hyperoxia." Invest Ophthalmol Vis Sci **45**(9): 3223-3228.

Cyprus Ministry of Agriculture Natural Resources and Environment. (1991). "1991 Census." Retrieved 1st February, 2010, from www.cyprus.gov.cy/moa/wdd/wdd.nsf/f8ca32f24081da0ac22573e000406fff/051403f075da65d5c22575cb00391be0/$FILE/ATTLE7GL/Report.pdf.

Cyprus Ministry of Finance. (2001). "2001 Census." Retrieved 1st February, 2010, from http://www.mof.gov.cy/mof/cystat/statistics.nsf/All/A5E0A9096615A568C2257122003FB72E/$file/POPULATION%20BY%20MUNICIPALITY_COMMUNITY-EN.xls?OpenElement.

D'Souza, Y., C. D. Short, et al. (2008). "Long-term follow-up of drusen-like lesions in patients with type II mesangiocapillary glomerulonephritis." Br J Ophthalmol **92**(7): 950-953.

Dalgaard, O. Z. (1957). "Bilateral polycystic disease of the kidneys; a follow-up of two hundred and eighty-four patients and their families." Acta Med Scand Suppl **328**: 1-255.

Davies, A. and B. P. Morgan (1993). "Expression of the glycosylphosphatidylinositol-linked complement-inhibiting protein CD59 antigen in insect cells using a baculovirus vector." Biochem J **295 (Pt 3)**: 889-896.

Davies, F., G. A. Coles, et al. (1991). "Polycystic kidney disease re-evaluated: a population-based study." Q J Med **79**(290): 477-485.

Davis, A. E., 3rd (2004). "Biological effects of C1 inhibitor." Drug News Perspect **17**(7): 439-446.

Davis, A. E., 3rd, R. A. Harrison, et al. (1984). "Physiologic inactivation of fluid phase C3b: isolation and structural analysis of C3c, C3d,g (alpha 2D), and C3g." J Immunol **132**(4): 1960-1966.

de Paula, P. F., J. E. Barbosa, et al. (2003). "Ontogeny of complement regulatory proteins - concentrations of factor h, factor I, c4b-binding protein, properdin and vitronectin in healthy children of different ages and in adults." Scand J Immunol **58**(5): 572-577.

Ding, K., M. Scortegagna, et al. (2005). "Retinal disease in mice lacking hypoxia-inducible transcription factor-2alpha." Invest Ophthalmol Vis Sci **46**(3): 1010-1016.

Dobo, J., V. Harmat, et al. (2009). "MASP-1, a promiscuous complement protease: structure of its catalytic region reveals the basis of its broad specificity." J Immunol **183**(2): 1207-1214.

Dorig, R. E., A. Marcil, et al. (1993). "The human CD46 molecule is a receptor for measles virus (Edmonston strain)." Cell **75**(2): 295-305.

Dragon-Durey, M. A., C. Blanc, et al. (2009). "The high frequency of complement factor H related CFHR1 gene deletion is restricted to specific subgroups of patients with atypical haemolytic uraemic syndrome." J Med Genet **46**(7): 447-450.

Druker, B. J., M. Talpaz, et al. (2001). "Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia." N Engl J Med **344**(14): 1031-1037.

Dumas, L., Y. H. Kim, et al. (2007). "Gene copy number variation spanning 60 million years of human and primate evolution." Genome Res **17**(9): 1266-1277.

Duvall-Young, J., M. K. MacDonald, et al. (1989). "Fundus changes in (type II) mesangiocapillary glomerulonephritis simulating drusen: a histopathological report." Br J Ophthalmol **73**(4): 297-302.

Ebert, B. L., J. D. Firth, et al. (1995). "Hypoxia and mitochondrial inhibitors regulate expression of glucose transporter-1 via distinct Cis-acting sequences." J Biol Chem **270**(49): 29083-29089.

Eckardt, K. U., S. T. Koury, et al. (1993). "Distribution of erythropoietin producing cells in rat kidneys during hypoxic hypoxia." Kidney Int **43**(4): 815-823.

Eckardt, K. U., A. Kurtz, et al. (1989). "Regulation of erythropoietin production is related to proximal tubular function." Am J Physiol **256**(5 Pt 2): F942-947.

Eckardt, K. U., C. W. Pugh, et al. (1994). "Production of erythropoietin by liver cells in vivo and in vitro." Ann N Y Acad Sci **718**: 50-60; discussion 61-53.

Eddy, C. A., J. M. MacCormick, et al. (2008). "Identification of large gene deletions and duplications in KCNQ1 and KCNH2 in patients with long QT syndrome." Heart Rhythm **5**(9): 1275-1281.

Elston, R. C. and J. Stewart (1971). "A general model for the genetic analysis of pedigree data." Hum Hered **21**(6): 523-542.

Epstein, A. C., J. M. Gleadle, et al. (2001). "C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation." Cell **107**(1): 43-54.

Escudier, B., T. Eisen, et al. (2007). "Sorafenib in advanced clear-cell renal-cell carcinoma." N Engl J Med **356**(2): 125-134.

Esparza-Gordillo, J., J. M. Soria, et al. (2004). "Genetic and environmental factors influencing the human factor H plasma levels." Immunogenetics **56**(2): 77-82.

European Polycystic Kidney Disease Consortium (1994). "The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. The European Polycystic Kidney Disease Consortium." Cell **78**(4): 725.

Fagan, K. A. and J. V. Weil (2001). "Potential genetic contributions to control of the pulmonary circulation and ventilation at high altitude." High Alt Med Biol **2**(2): 165-171.

Ferreira, V. P., A. P. Herbert, et al. (2006). "Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces." J Immunol **177**(9): 6308-6316.

Firth, J. D., B. L. Ebert, et al. (1995). "Hypoxic regulation of lactate dehydrogenase A. Interaction between hypoxia-inducible factor 1 and cAMP response elements." J Biol Chem **270**(36): 21021-21027.

Folkman, J., K. Watson, et al. (1989). "Induction of angiogenesis during the transition from hyperplasia to neoplasia." Nature **339**(6219): 58-61.

Fong, G. H. (2008). "Mechanisms of adaptive angiogenesis to tissue hypoxia." Angiogenesis **11**(2): 121-140.

Freedman, B. I., N. V. Volkova, et al. (2005). "Population-based screening for family history of end-stage renal disease among incident dialysis patients." Am J Nephrol **25**(6): 529-535.

Fremeaux-Bacchi, V., M. A. Dragon-Durey, et al. (2004). "Complement factor I: a susceptibility gene for atypical haemolytic uraemic syndrome." J Med Genet **41**(6): e84.

Fremeaux-Bacchi, V., E. C. Miller, et al. (2008). "Mutations in complement C3 predispose to development of atypical hemolytic uremic syndrome." Blood **112**(13): 4948-4952.

Fukuda, R., H. Zhang, et al. (2007). "HIF-1 regulates cytochrome oxidase subunits to optimize efficiency of respiration in hypoxic cells." Cell **129**(1): 111-122.

Gable, M., M. Williams, et al. (2003). "Comparative multiplex dosage analysis detects whole exon deletions at the phenylalanine hydroxylase locus." Hum Mutat **21**(4): 379-386.

Gaggar, A., D. M. Shayakhmetov, et al. (2003). "CD46 is a cellular receptor for group B adenoviruses." Nat Med **9**(11): 1408-1412.

Gale, D. P. and P. H. Maxwell (2010). "The role of HIF in immunity." Int J Biochem Cell Biol **42**(4): 486-494.

Garred, P., M. Harboe, et al. (1994). "Dual role of mannan-binding protein in infections: another case of heterosis?" Eur J Immunogenet **21**(2): 125-131.

Garruto, R. M., C. T. Chin, et al. (2003). "Hematological differences during growth among Tibetans and Han Chinese born and raised at high altitude in Qinghai, China." Am J Phys Anthropol **122**(2): 171-183.

Genel, F., A. G. Sjoholm, et al. (2005). "Complement factor I deficiency associated with recurrent infections, vasculitis and immune complex glomerulonephritis." Scand J Infect Dis **37**(8): 615-618.

GENEMILL (2003). Birth of Aphrodite and the origin of maternal lineages in Cyprus. 3rd Annual CoE Meeting. Paphos, Cyprus**:** Table 3.

Genin, E., A. Tullio-Pelet, et al. (2004). "Estimating the age of rare disease mutations: the example of Triple-A syndrome." J Med Genet **41**(6): 445-449.

Gibson, J. S., P. I. Milner, et al. (2008). "Oxygen and reactive oxygen species in articular cartilage: modulators of ionic homeostasis." Pflugers Arch **455**(4): 563-573.

Gilbert, J. S., M. J. Ryan, et al. (2008). "Pathophysiology of hypertension during preeclampsia: linking placental ischemia with endothelial dysfunction." Am J Physiol Heart Circ Physiol **294**(2): H541-550.

Giussani, D. A. (2007). "Hypoxia, fetal growth and early origins of disease: the Andean curse on the Conquistadors." J Physiol **582**(Pt 2): 472.

Gnarra, J. R., K. Tory, et al. (1994). "Mutations of the VHL tumour suppressor gene in renal carcinoma." Nat Genet **7**(1): 85-90.

Goicoechea de Jorge, E., C. L. Harris, et al. (2007). "Gain-of-function mutations in complement factor B are associated with atypical hemolytic uremic syndrome." Proc Natl Acad Sci U S A **104**(1): 240-245.

Goldberg, M. A., G. A. Glass, et al. (1987). "The regulated expression of erythropoietin by two human hepatoma cell lines." Proc Natl Acad Sci U S A **84**(22): 7972-7976.

Gonzales, G. F. (2007). "Peruvian contributions to the study on human reproduction at high altitude: from the chronicles of the Spanish conquest to the present." Respir Physiol Neurobiol **158**(2-3): 172-179.

Goodman, J. (1947). "Production of polycythemia in rabbits by anoxia and cobalt." Proc Soc Exp Biol Med **64**(3): 336-342.

Goodship, T. H. and D. Kavanagh (2010). "Pulling the trigger in atypical hemolytic uremic syndrome: the role of pregnancy." J Am Soc Nephrol **21**(5): 731-732.

Gordan, J. D., P. Lal, et al. (2008). "HIF-alpha effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma." Cancer Cell **14**(6): 435-446.

Gordan, J. D. and M. C. Simon (2007). "Hypoxia-inducible factors: central regulators of the tumor phenotype." Curr Opin Genet Dev **17**(1): 71-77.

Groves, B. M., J. T. Reeves, et al. (1987). "Operation Everest II: elevated high-altitude pulmonary resistance unresponsive to oxygen." J Appl Physiol **63**(2): 521-530.

Gruber, M., C. J. Hu, et al. (2007). "Acute postnatal ablation of Hif-2alpha results in anemia." Proc Natl Acad Sci U S A **104**(7): 2301-2306.

Gruber, S. B., M. M. Entius, et al. (1998). "Pathogenesis of adenocarcinoma in Peutz-Jeghers syndrome." Cancer Res **58**(23): 5267-5270.

Gudbjartsson, D. F., K. Jonasson, et al. (2000). "Allegro, a new computer program for multipoint linkage analysis." Nat Genet **25**(1): 12-13.

Gupta, M. L., K. S. Rao, et al. (1992). "Lack of smooth muscle in the small pulmonary arteries of the native Ladakhi. Is the Himalayan highlander adapted?" Am Rev Respir Dis **145**(5): 1201-1204.

Gusfield, D. (2001). "Inference of haplotypes from samples of diploid populations: complexity and algorithms." J Comput Biol **8**(3): 305-323.

Hageman, G. S., D. H. Anderson, et al. (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**(20): 7227-7232.

Hajela, K., M. Kojima, et al. (2002). "The biological functions of MBL-associated serine proteases (MASPs)." Immunobiology **205**(4-5): 467-475.

Halpin, D. (2008). Chronic kidney disease: national clinical guideline for early identification and management in adults in primary and secondary care. N. C. C. f. C. Conditions, Royal College of Physicians, London**:** 147-148.

Hansard (1960). HL Deb. **223:** 929-934.

Hastings, P. J., J. R. Lupski, et al. (2009). "Mechanisms of change in gene copy number." Nat Rev Genet **10**(8): 551-564.

Heinen, S., A. Hartmann, et al. (2009). "Factor H-related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation." Blood **114**(12): 2439-2447.

Heinen, S., P. Sanchez-Corral, et al. (2006). "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**(3): 292-293.

Helmy, K. Y., K. J. Katschke, Jr., et al. (2006). "CRIg: a macrophage complement receptor required for phagocytosis of circulating pathogens." Cell **124**(5): 915-927.

Hewitson, K. S., L. A. McNeill, et al. (2002). "Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family." J Biol Chem **277**(29): 26351-26355.

Hickey, M. M., J. C. Lam, et al. (2007). "von Hippel-Lindau mutation in mice recapitulates Chuvash polycythemia via hypoxia-inducible factor-2alpha signaling and splenic erythropoiesis." J Clin Invest **117**(12): 3879-3889.

Hickey, M. M., T. Richardson, et al. (2010). "The von Hippel-Lindau Chuvash mutation promotes pulmonary hypertension and fibrosis in mice." J Clin Invest **120**(3): 827-839.

Hirsila, M., P. Koivunen, et al. (2003). "Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor." J Biol Chem **278**(33): 30772-30780.

Hoeprich, P. D., Jr., C. A. Dahinden, et al. (1985). "A synthetic nonapeptide corresponding to the NH2-terminal sequence of C3d-K causes leukocytosis in rabbits." J Biol Chem **260**(5): 2597-2600.

Hoffmann, K. and T. H. Lindner (2005). "easyLINKAGE-Plus--automated linkage analyses using large-scale SNP data." Bioinformatics **21**(17): 3565-3567.

Hom, G., R. R. Graham, et al. (2008). "Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX." N Engl J Med **358**(9): 900-909.

Howarth, C., I. Chanarin, et al. (1979). "Familial erythrocytosis." Scand J Haematol **23**(3): 217-222.

Hu, C. J., A. Sataur, et al. (2007). "The N-terminal transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1alpha and HIF-2alpha." Mol Biol Cell **18**(11): 4528-4542.

Hu, X., S. D. Chakravarty, et al. (2008). "Regulation of interferon and Toll-like receptor signaling during macrophage activation by opposing feedforward and feedback inhibition mechanisms." Immunol Rev **226**: 41-56.

Illumina. (2010). "CytoSNP 12 Datasheet." Retrieved February 7th, 2010, from http://www.illumina.com.

Imagawa, S., M. A. Goldberg, et al. (1991). "Regulatory elements of the erythropoietin gene." Blood **77**(2): 278-285.

Irwin, J., J. Saunier, et al. (2008). "Mitochondrial control region sequences from northern Greece and Greek Cypriots." Int J Legal Med **122**(1): 87-89.

Jacobs, K., C. Shoemaker, et al. (1985). "Isolation and characterization of genomic and cDNA clones of human erythropoietin." Nature **313**(6005): 806-810.

Jacobson, L. O., E. Goldwasser, et al. (1957). "Role of the kidney in erythropoiesis." Nature **179**(4560): 633-634.

Jeanette, J. C. (2008). Renal Pathology. Rubin's Pathology. Philadelphia, Lipincott Williams and Wilkins**:** 714.

Jemal, A., R. Siegel, et al. (2009). "Cancer statistics, 2009." CA Cancer J Clin **59**(4): 225-249.

Johnson, U., K. Ohlsson, et al. (1976). "Effects of granulocyte neutral proteases on complement components." Scand J Immunol **5**(4): 421-426.

Jozsi, M. and P. F. Zipfel (2008). "Factor H family proteins and human diseases." Trends Immunol **29**(8): 380-387.

Julian, C. G., H. L. Galan, et al. (2008). "Lower uterine artery blood flow and higher endothelin relative to nitric oxide metabolite levels are associated with reductions in birth weight at high altitude." Am J Physiol Regul Integr Comp Physiol **295**(3): R906-915.

Kaelin, W. G., Jr. (2008). "The von Hippel-Lindau tumour suppressor protein: O2 sensing and cancer." Nat Rev Cancer **8**(11): 865-873.

Kaelin, W. G., Jr. and P. J. Ratcliffe (2008). "Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway." Mol Cell **30**(4): 393-402.

Kallstrom, H., D. Blackmer Gill, et al. (2001). "Attachment of Neisseria gonorrhoeae to the cellular pilus receptor CD46: identification of domains important for bacterial adherence." Cell Microbiol **3**(3): 133-143.

Kawai, T., S. Takeshita, et al. (2009). "Associations between decay-accelerating factor polymorphisms and allergic respiratory diseases." Clin Exp Allergy **39**(10): 1508-1514.

Kazatchkine, M. D., D. T. Fearon, et al. (1979). "Human alternative complement pathway: membrane-associated sialic acid regulates the competition between B and beta1 H for cell-bound C3b." J Immunol **122**(1): 75-81.

Kemper, C., A. C. Chan, et al. (2003). "Activation of human CD4+ cells with CD3 and CD46 induces a T-regulatory cell 1 phenotype." Nature **421**(6921): 388-392.

Keyes, L. E., J. F. Armaza, et al. (2003). "Intrauterine growth restriction, preeclampsia, and intrauterine mortality at high altitude in Bolivia." Pediatr Res **54**(1): 20-25.

Kim, D. D. and W. C. Song (2006). "Membrane complement regulatory proteins." Clin Immunol **118**(2-3): 127-136.

Kim, J. S., T. B. Krasieva, et al. (2002). "Specific recruitment of human cohesin to laser-induced DNA damage." J Biol Chem **277**(47): 45149-45153.

Kim, J. W., I. Tchernyshyov, et al. (2006). "HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia." Cell Metab **3**(3): 177-185.

Klos, A., A. J. Tenner, et al. (2009). "The role of the anaphylatoxins in health and disease." Mol Immunol **46**(14): 2753-2766.

Knapp, A. B., V. Kassianidou, et al. (2001). "Copper Smelting in Late Bronze Age Cyprus: The Excavations at Politiko Phorades." Near Eastern Archaeology **64**(4): 204-210.

Koivunen, P., M. Hirsila, et al. (2004). "Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases." J Biol Chem **279**(11): 9899-9904.

Kondo, K., J. Klco, et al. (2002). "Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein." Cancer Cell **1**(3): 237-246.

Korbel, J. O., A. E. Urban, et al. (2007). "Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome." Proc Natl Acad Sci U S A **104**(24): 10110-10115.

Kosambi, D. D. (1944). "The estimation of map distances from recombination values." Annals of Eugenics **12**: 172-175.

Koury, S. T., M. C. Bondurant, et al. (1991). "Localization of cells producing erythropoietin in murine liver by in situ hybridization." Blood **77**(11): 2497-2503.

Kruglyak, L., M. J. Daly, et al. (1996). "Parametric and nonparametric linkage analysis: a unified multipoint approach." Am J Hum Genet **58**(6): 1347-1363.

Kyrri, A. R., X. Felekis, et al. (2009). "Hemoglobin variants in Cyprus." Hemoglobin **33**(2): 81-94.

Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." J Mol Biol **157**(1): 105-132.

Lachmann, P. J. (2009). "The amplification loop of the complement pathways." Adv Immunol **104**: 115-149.

Lagueux, M., E. Perrodou, et al. (2000). "Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of Drosophila." Proc Natl Acad Sci U S A **97**(21): 11427-11432.

Lander, E. S. and P. Green (1987). "Construction of multilocus genetic linkage maps in humans." Proc Natl Acad Sci U S A **84**(8): 2363-2367.

Lando, D., D. J. Peet, et al. (2002). "FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor." Genes Dev **16**(12): 1466-1471.

Lathrop, G. M., J. M. Lalouel, et al. (1984). "Strategies for multilocus linkage analysis in humans." Proc Natl Acad Sci U S A **81**(11): 3443-3446.

Laughner, E., P. Taghavi, et al. (2001). "HER2 (neu) signaling increases the rate of hypoxia-inducible factor 1alpha (HIF-1alpha) synthesis: novel mechanism for HIF-1-mediated vascular endothelial growth factor expression." Mol Cell Biol **21**(12): 3995-4004.

Lee, H. Y., A. Andalibi, et al. (2004). "Antimicrobial activity of innate immune molecules against Streptococcus pneumoniae, Moraxella catarrhalis and nontypeable Haemophilus influenzae." BMC Infect Dis **4**: 12.

Lee, J. A., K. Inoue, et al. (2006). "Role of genomic architecture in PLP1 duplication causing Pelizaeus-Merzbacher disease." Hum Mol Genet **15**(14): 2250-2265.

Levashina, E. A., L. F. Moita, et al. (2001). "Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, Anopheles gambiae." Cell **104**(5): 709-718.

Lewis, M. A. (2008). "Demography of renal disease in childhood." Semin Fetal Neonatal Med **13**(3): 118-124.

Licht, C. and V. Fremeaux-Bacchi (2009). "Hereditary and acquired complement dysregulation in membranoproliferative glomerulonephritis." Thromb Haemost **101**(2): 271-278.

Lieber, M. R. (2008). "The mechanism of human nonhomologous DNA end joining." J Biol Chem **283**(1): 1-5.

Liskay, R. M., A. Letsou, et al. (1987). "Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells." Genetics **115**(1): 161-167.

Liu, Y. L., S. O. Ang, et al. (2004). "Regulation of ferrochelatase gene expression by hypoxia." Life Sci **75**(17): 2035-2043.

Loizidou, M. A., T. Michael, et al. (2008). "Genetic polymorphisms in the DNA repair genes XRCC1, XRCC2 and XRCC3 and risk of breast cancer in Cyprus." Breast Cancer Res Treat **112**(3): 575-579.

Lublin, D. M., M. K. Liszewski, et al. (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**(1): 181-194.

Lum, J. J., T. Bui, et al. (2007). "The transcription factor HIF-1alpha plays a critical role in the growth factor-dependent regulation of both aerobic and anaerobic glycolysis." Genes Dev **21**(9): 1037-1049.

Lupski, J. R. (1998). "Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits." Trends Genet **14**(10): 417-422.

Madan, A. and P. T. Curtin (1993). "A 24-base-pair sequence 3' to the human erythropoietin gene contains a hypoxia-responsive transcriptional enhancer." Proc Natl Acad Sci U S A **90**(9): 3928-3932.

Makino, Y., R. Cao, et al. (2001). "Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression." Nature **414**(6863): 550-554.

Marketos, S. G., A. G. Eftychiadis, et al. (1993). "Acute renal failure according to ancient Greek and Byzantine medical writers." J R Soc Med **86**(5): 290-293.

Markianos, K., M. J. Daly, et al. (2001). "Efficient multipoint linkage analysis through reduction of inheritance space." Am J Hum Genet **68**(4): 963-977.

Matsumoto, A. K., J. Kopicky-Burd, et al. (1991). "Intersection of the complement and immune systems: a signal transduction complex of the B lymphocyte-containing complement receptor type 2 and CD19." J Exp Med **173**(1): 55-64.

Maxwell, P. H., G. U. Dachs, et al. (1997). "Hypoxia-inducible factor-1 modulates gene expression in solid tumors and influences both angiogenesis and tumor growth." Proc Natl Acad Sci U S A **94**(15): 8104-8109.

Maxwell, P. H., D. J. Ferguson, et al. (1994). "Expression of a homologously recombined erythopoietin-SV40 T antigen fusion gene in mouse liver: evidence for erythropoietin production by Ito cells." Blood **84**(6): 1823-1830.

Maxwell, P. H., M. K. Osmond, et al. (1993). "Identification of the renal erythropoietin-producing cells using transgenic mice." Kidney Int **44**(5): 1149-1162.

Maxwell, P. H., C. W. Pugh, et al. (1993). "Inducible operation of the erythropoietin 3' enhancer in multiple cell lines: evidence for a widespread oxygen-sensing mechanism." Proc Natl Acad Sci U S A **90**(6): 2423-2427.

Maxwell, P. H., M. S. Wiesener, et al. (1999). "The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis." Nature **399**(6733): 271-275.

Maynard, M. A., A. J. Evans, et al. (2007). "Dominant-negative HIF-3 alpha 4 suppresses VHL-null renal cell carcinoma progression." Cell Cycle **6**(22): 2810-2816.

McCarroll, S. A., F. G. Kuruvilla, et al. (2008). "Integrated detection and population-genetic analysis of SNPs and copy number variation." Nat Genet **40**(10): 1166-1174.

McDonough, M. A., V. Li, et al. (2006). "Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2)." Proc Natl Acad Sci U S A **103**(26): 9814-9819.

McRae, J. L., P. J. Cowan, et al. (2001). "Human factor H-related protein 5 (FHR-5). A new complement-associated protein." J Biol Chem **276**(9): 6747-6754.

McRae, J. L., T. G. Duthy, et al. (2005). "Human factor H-related protein 5 has cofactor activity, inhibits C3 convertase activity, binds heparin and C-reactive protein, and associates with lipoprotein." J Immunol **174**(10): 6250-6256.

McVey, M. and S. E. Lee (2008). "MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings." Trends Genet **24**(11): 529-538.

Meri, S. and M. K. Pangburn (1990). "Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding site on factor H." Proc Natl Acad Sci U S A **87**(10): 3982-3986.

Meri, S., H. Waldmann, et al. (1991). "Distribution of protectin (CD59), a complement membrane attack inhibitor, in normal human tissues." Lab Invest **65**(5): 532-537.

Metzen, E., U. Berchner-Pfannschmidt, et al. (2003). "Intracellular localisation of human HIF-1 alpha hydroxylases: implications for oxygen sensing." J Cell Sci **116**(Pt 7): 1319-1326.

Meuth, J. L., E. L. Morgan, et al. (1983). "Suppression of T lymphocyte functions by human C3 fragments. I. Inhibition of human T cell proliferative responses by a kallikrein cleavage fragment of human iC3b." J Immunol **130**(6): 2605-2611.

Miwa, T. and W. C. Song (2001). "Membrane complement regulatory proteins: insight from animal studies and relevance to human diseases." Int Immunopharmacol **1**(3): 445-459.

Miyake, T., C. K. Kung, et al. (1977). "Purification of human erythropoietin." J Biol Chem **252**(15): 5558-5564.

Monge, C. (1943). "Chronic Mountain Sickness." Physiological Review **23**: 166-188.

Monteferrante, G., S. Brioschi, et al. (2007). "Genetic analysis of the complement factor H related 5 gene in haemolytic uraemic syndrome." Mol Immunol **44**(7): 1704-1708.

Montes, T., E. Goicoechea de Jorge, et al. (2008). "Genetic deficiency of complement factor H in a patient with age-related macular degeneration and membranoproliferative glomerulonephritis." Mol Immunol **45**(10): 2897-2904.

Moore, I., L. Strain, et al. (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**(2): 379-387.

Moore, L. G. (2001). "Human genetic adaptation to high altitude." High Alt Med Biol **2**(2): 257-279.

Moore, L. G., S. Niermeyer, et al. (1998). "Human adaptation to high altitude: regional and life-cycle perspectives." Am J Phys Anthropol **Suppl 27**: 25-64.

Moore, L. G., M. Shriver, et al. (2004). "Maternal adaptation to high-altitude pregnancy: an experiment of nature--a review." Placenta **25 Suppl A**: S60-71.

Morita, M., O. Ohneda, et al. (2003). "HLF/HIF-2alpha is a key factor in retinopathy of prematurity in association with erythropoietin." EMBO J **22**(5): 1134-1146.

Motzer, R. J., T. E. Hutson, et al. (2007). "Sunitinib versus interferon alfa in metastatic renal-cell carcinoma." N Engl J Med **356**(2): 115-124.

Moura, S. B., L. R. Almeida, et al. (2008). "Toll-like receptor (TLR2, TLR4 and TLR5) gene polymorphisms and Helicobacter pylori infection in children with and without duodenal ulcer." Microbes Infect **10**(14-15): 1477-1483.

Murphy, B., T. Georgiou, et al. (2002). "Factor H-related protein-5: a novel component of human glomerular immune deposits." Am J Kidney Dis **39**(1): 24-27.

Nakajima, T., E. E. Nakayama, et al. (2009). "Impact of novel TRIM5alpha variants, Gly110Arg and G176del, on the anti-HIV-1 activity and the susceptibility to HIV-1 infection." AIDS **23**(16): 2091-2100.

Narendra, U., G. J. Pauer, et al. (2009). "Genetic analysis of complement factor H related 5, CFHR5, in patients with age-related macular degeneration." Mol Vis **15**: 731-736.

Necas, E. and E. B. Thorling (1972). "Unresponsiveness of erythropoietin-producing cells to cyanide." Am J Physiol **222**(5): 1187-1190.

Newcombe, R. G. (1998). "Two-sided confidence intervals for the single proportion: comparison of seven methods." Stat Med **17**(8): 857-872.

Nguyen, D. Q., C. Webber, et al. (2008). "Reduced purifying selection prevails over positive selection in human copy number variant evolution." Genome Res **18**(11): 1711-1723.

Niermeyer, S., P. Yang, et al. (1995). "Arterial oxygen saturation in Tibetan and Han infants born in Lhasa, Tibet." N Engl J Med **333**(19): 1248-1252.

Nonaka, M. and A. Kimura (2006). "Genomic view of the evolution of the complement system." Immunogenetics **58**(9): 701-713.

Noris, M., S. Brioschi, et al. (2003). "Familial haemolytic uraemic syndrome and an MCP mutation." Lancet **362**(9395): 1542-1547.

Noris, M., S. Bucchioni, et al. (2005). "Complement factor H mutation in familial thrombotic thrombocytopenic purpura with ADAMTS13 deficiency and renal involvement." J Am Soc Nephrol **16**(5): 1177-1183.

Novembre, J., T. Johnson, et al. (2008). "Genes mirror geography within Europe." Nature **456**(7218): 98-101.

Nytko, K. J., P. Spielmann, et al. (2007). "Regulated function of the prolyl-4-hydroxylase domain (PHD) oxygen sensor proteins." Antioxid Redox Signal **9**(9): 1329-1338.

O'Connell, J. R. and D. E. Weeks (1995). "The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance." Nat Genet **11**(4): 402-408.

O'Connell, J. R. and D. E. Weeks (1998). "PedCheck: a program for identification of genotype incompatibilities in linkage analysis." Am J Hum Genet **63**(1): 259-266.

O'Rourke, J. F., G. U. Dachs, et al. (1997). "Hypoxia response elements." Oncol Res **9**(6-7): 327-332.

Ott, J. (1974). "Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies." Am J Hum Genet **26**(5): 588-597.

Pan, Y., K. D. Mansfield, et al. (2007). "Multiple factors affecting cellular redox status and energy metabolism modulate hypoxia-inducible factor prolyl hydroxylase activity in vivo and in vitro." Mol Cell Biol **27**(3): 912-925.

Pangburn, M. K., D. C. Morrison, et al. (1980). "Activation of the alternative complement pathway: recognition of surface structures on activators by bound C3b." J Immunol **124**(2): 977-982.

Pangburn, M. K., R. D. Schreiber, et al. (1977). "Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein beta1H for cleavage of C3b and C4b in solution." J Exp Med **146**(1): 257-270.

Pangburn, M. K., R. D. Schreiber, et al. (1981). "Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3." J Exp Med **154**(3): 856-867.

Papandreou, I., R. A. Cairns, et al. (2006). "HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption." Cell Metab **3**(3): 187-197.

Parker, C. (2009). "Eculizumab for paroxysmal nocturnal haemoglobinuria." Lancet **373**(9665): 759-767.

Pei, S. X., X. J. Chen, et al. (1989). "Chronic mountain sickness in Tibet." Q J Med **71**(266): 555-574.

Penaloza, D. and J. Arias-Stella (2007). "The heart and pulmonary circulation at high altitudes: healthy highlanders and chronic mountain sickness." Circulation **115**(9): 1132-1146.

Peng, Q., K. Li, et al. (2009). "The role of anaphylatoxins C3a and C5a in regulating innate and adaptive immune responses." Inflamm Allergy Drug Targets **8**(3): 236-246.

Peng, X. H., P. Karna, et al. (2006). "Cross-talk between epidermal growth factor receptor and hypoxia-inducible factor-1alpha signal pathways increases resistance to apoptosis by up-regulating survivin gene expression." J Biol Chem **281**(36): 25903-25914.

Percy, M. J., P. A. Beer, et al. (2008). "Novel Exon 12 mutations in the HIF2A gene associated with erythrocytosis." Blood.

Percy, M. J., P. W. Furlow, et al. (2007). "A novel erythrocytosis-associated PHD2 mutation suggests the location of a HIF binding groove." Blood **110**(6): 2193-2196.

Percy, M. J., P. W. Furlow, et al. (2008). "A gain-of-function mutation in the HIF2A gene in familial erythrocytosis." N Engl J Med **358**(2): 162-168.

Percy, M. J., Q. Zhao, et al. (2006). "A family with erythrocytosis establishes a role for prolyl hydroxylase domain protein 2 in oxygen homeostasis." Proc Natl Acad Sci U S A **103**(3): 654-659.

Perry, G. H., N. J. Dominy, et al. (2007). "Diet and the evolution of human amylase gene copy number variation." Nat Genet **39**(10): 1256-1260.

Pickering, M. C. and H. T. Cook (2008). "Translational mini-review series on complement factor H: renal diseases associated with complement factor H: novel insights from humans and animals." Clin Exp Immunol **151**(2): 210-230.

Pickering, M. C., H. T. Cook, et al. (2002). "Uncontrolled C3 activation causes membranoproliferative glomerulonephritis in mice deficient in complement factor H." Nat Genet **31**(4): 424-428.

Pickering, M. C., E. G. de Jorge, et al. (2007). "Spontaneous hemolytic uremic syndrome triggered by complement factor H lacking surface recognition domains." J Exp Med **204**(6): 1249-1256.

Pinto, M. R., D. Melillo, et al. (2007). "Ancient origin of the complement system: emerging invertebrate models." Adv Exp Med Biol **598**: 372-388.

Poliakova, L. A. (1974). "[Familial erythrocytosis among the residents of the Chuvash ASSR]." Probl Gematol Pereliv Krovi **19**(10): 30-33.

Pollard, P. J., J. J. Briere, et al. (2005). "Accumulation of Krebs cycle intermediates and over-expression of HIF1alpha in tumours which result from germline FH and SDH mutations." Hum Mol Genet **14**(15): 2231-2239.

Pollard, P. J., B. Spencer-Dene, et al. (2007). "Targeted inactivation of fh1 causes proliferative renal cyst development and activation of the hypoxia pathway." Cancer Cell **11**(4): 311-319.

Preston, I. R. (2007). "Clinical perspective of hypoxia-mediated pulmonary hypertension." Antioxid Redox Signal **9**(6): 711-721.

Pugh, C. W. and P. J. Ratcliffe (2003). "Regulation of angiogenesis by hypoxia: role of the HIF system." Nat Med **9**(6): 677-684.

Pugh, C. W., C. C. Tan, et al. (1991). "Functional analysis of an oxygen-regulated transcriptional enhancer lying 3' to the mouse erythropoietin gene." Proc Natl Acad Sci U S A **88**(23): 10553-10557.

Pugh, L. G. (1964). "Blood Volume and Haemoglobin Concentration at Altitudes above 18,000 Ft. (5500 M)." J Physiol **170**: 344-354.

Rankin, E. B., M. P. Biju, et al. (2007). "Hypoxia-inducible factor-2 (HIF-2) regulates hepatic erythropoietin in vivo." J Clin Invest **117**(4): 1068-1077.

Ratcliffe, P. J. (2007). "Fumarate hydratase deficiency and cancer: activation of hypoxia signaling?" Cancer Cell **11**(4): 303-305.

Rayssiguier, C., D. S. Thaler, et al. (1989). "The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants." Nature **342**(6248): 396-401.

Reich, D. E., M. Cargill, et al. (2001). "Linkage disequilibrium in the human genome." Nature **411**(6834): 199-204.

Reiche, E. M., M. A. Ehara Watanabe, et al. (2008). "Frequency of CCR5-Delta32 deletion in human immunodeficiency virus type 1 (HIV-1) in healthy blood donors, HIV-1-exposed seronegative and HIV-1-seropositive individuals of southern Brazilian population." Int J Mol Med **22**(5): 669-675.

Richards, A., E. J. Kemp, et al. (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**(22): 12966-12971.

Risitano, A. M., R. Notaro, et al. (2009). "Complement fraction 3 binding on erythrocytes as additional mechanism of disease in paroxysmal nocturnal hemoglobinuria patients treated by eculizumab." Blood **113**(17): 4094-4100.

Rodriguez de Cordoba, S., J. Esparza-Gordillo, et al. (2004). "The human complement factor H: functional roles, genetic variations and disease associations." Mol Immunol **41**(4): 355-367.

Rodriguez de Cordoba, S., D. M. Lublin, et al. (1985). "Human genes for three complement components that regulate the activation of C3 are tightly linked." J Exp Med **161**(5): 1189-1195.

Rose, K. L., D. Paixao-Cavalcante, et al. (2008). "Factor I is required for the development of membranoproliferative glomerulonephritis in factor H-deficient mice." J Clin Invest **118**(2): 608-618.

Rosenberger, C., S. Mandriota, et al. (2002). "Expression of hypoxia-inducible factor-1alpha and -2alpha in hypoxic and ischemic rat kidneys." J Am Soc Nephrol **13**(7): 1721-1732.

Ross, G. D., J. D. Lambris, et al. (1982). "Generation of three different fragments of bound C3 with purified factor I or serum. I. Requirements for factor H vs CR1 cofactor activity." J Immunol **129**(5): 2051-2060.

Rowe, J. A., J. M. Moulds, et al. (1997). "P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1." Nature **388**(6639): 292-295.

Rupert, J. L. and M. S. Koehle (2006). "Evidence for a genetic basis for altitude-related illness." High Alt Med Biol **7**(2): 150-167.

Sabeti, P. C., P. Varilly, et al. (2007). "Genome-wide detection and characterization of positive selection in human populations." Nature **449**(7164): 913-918.

Sadallah, S., F. Gudat, et al. (1999). "Glomerulonephritis in a patient with complement factor I deficiency." Am J Kidney Dis **33**(6): 1153-1157.

Sanchez-Corral, P., C. Gonzalez-Rubio, et al. (2004). "Functional analysis in serum from atypical Hemolytic Uremic Syndrome patients reveals impaired protection of host cells associated with mutations in factor H." Mol Immunol **41**(1): 81-84.

Sanchez-Corral, P., D. Perez-Caballero, et al. (2002). "Structural and functional characterization of factor H mutations associated with atypical hemolytic uremic syndrome." Am J Hum Genet **71**(6): 1285-1295.

Santoro, F., P. E. Kennedy, et al. (1999). "CD46 is a cellular receptor for human herpesvirus 6." Cell **99**(7): 817-827.

Satko, S. G., J. R. Sedor, et al. (2007). "Familial clustering of chronic kidney disease." Semin Dial **20**(3): 229-236.

Schaschl, H., T. J. Aitman, et al. (2009). "Copy number variation in the human genome and its implication in autoimmunity." Clin Exp Immunol **156**(1): 12-16.

Schiffmann, R., J. B. Kopp, et al. (2001). "Enzyme replacement therapy in Fabry disease: a randomized controlled trial." JAMA **285**(21): 2743-2749.

Schmid-Kubista, K. E., N. Tosakulwong, et al. (2009). "Contribution of copy number variation in the regulation of complement activation locus to development of age-related macular degeneration." Invest Ophthalmol Vis Sci **50**(11): 5070-5079.

Schmidt, C. Q., A. P. Herbert, et al. (2008). "Translational mini-review series on complement factor H: structural and functional correlations for factor H." Clin Exp Immunol **151**(1): 14-24.

Schofield, C. J. and P. J. Ratcliffe (2004). "Oxygen sensing by HIF hydroxylases." Nat Rev Mol Cell Biol **5**(5): 343-354.

Schorey, J. S., M. C. Carroll, et al. (1997). "A macrophage invasion mechanism of pathogenic mycobacteria." Science **277**(5329): 1091-1093.

Scortegagna, M., K. Ding, et al. (2005). "HIF-2alpha regulates murine hematopoietic development in an erythropoietin-dependent manner." Blood **105**(8): 3133-3140.

Semenza, G. L., R. C. Dureza, et al. (1990). "Human erythropoietin gene expression in transgenic mice: multiple transcription initiation sites and cis-acting regulatory elements." Mol Cell Biol **10**(3): 930-938.

Semenza, G. L., P. H. Roth, et al. (1994). "Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1." J Biol Chem **269**(38): 23757-23763.

Serganova, I., J. Humm, et al. (2006). "Tumor hypoxia imaging." Clin Cancer Res **12**(18): 5260-5264.

Servais, A., V. Fremeaux-Bacchi, et al. (2007). "Primary glomerulonephritis with isolated C3 deposits: a new entity which shares common genetic risk factors with haemolytic uraemic syndrome." J Med Genet **44**(3): 193-199.

Simonson, T. S., Y. Yang, et al. (2010). "Genetic Evidence for High-Altitude Adaptation in Tibet." Science.

Skerka, C. and P. F. Zipfel (2008). "Complement factor H related proteins in immune diseases." Vaccine **26 Suppl 8**: I9-14.

Smith, B. O., R. L. Mallin, et al. (2002). "Structure of the C3b binding site of CR1 (CD35), the immune adherence receptor." Cell **108**(6): 769-780.

Smith, T. G., J. T. Brooks, et al. (2006). "Mutation of von Hippel-Lindau tumour suppressor and human cardiopulmonary physiology." PLoS Med **3**(7): e290.

Sobel, E. and K. Lange (1996). "Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics." Am J Hum Genet **58**(6): 1323-1337.

Solovjov, D. A., E. Pluskota, et al. (2005). "Distinct roles for the alpha and beta subunits in the functions of integrin alphaMbeta2." J Biol Chem **280**(2): 1336-1345.

Sowter, H. M., R. R. Raval, et al. (2003). "Predominant role of hypoxia-inducible transcription factor (Hif)-1alpha versus Hif-2alpha in regulation of the transcriptional response to hypoxia." Cancer Res **63**(19): 6130-6134.

Stolze, I. P., Y. M. Tian, et al. (2004). "Genetic analysis of the role of the asparaginyl hydroxylase factor inhibiting hypoxia-inducible factor (HIF) in regulating HIF transcriptional target genes." J Biol Chem **279**(41): 42719-42725.

Stuber, T., C. Sartori, et al. (2008). "Respiratory nitric oxide and pulmonary artery pressure in Aymara and Caucasian children at high altitude." Chest.

Su, A. I., T. Wiltshire, et al. (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes." Proc Natl Acad Sci U S A **101**(16): 6062-6067.

Summerfield, J. A., M. Sumiya, et al. (1997). "Association of mutations in mannose binding protein gene with childhood infection in consecutive hospital series." BMJ **314**(7089): 1229-1232.

Takeda, K., A. Cowan, et al. (2007). "Essential role for prolyl hydroxylase domain protein 2 in oxygen homeostasis of the adult vascular system." Circulation **116**(7): 774-781.

Tan, C. C. and P. J. Ratcliffe (1991). "Effect of inhibitors of oxidative phosphorylation on erythropoietin mRNA in isolated perfused rat kidneys." Am J Physiol **261**(6 Pt 2): F982-987.

Thiele, H. and P. Nurnberg (2005). "HaploPainter: a tool for drawing pedigrees with complex haplotypes." Bioinformatics **21**(8): 1730-1732.

Treins, C., S. Giorgetti-Peraldi, et al. (2005). "Regulation of hypoxia-inducible factor (HIF)-1 activity and expression of HIF hydroxylases in response to insulin-like growth factor I." Mol Endocrinol **19**(5): 1304-1317.

Tuzun, E., J. A. Bailey, et al. (2004). "Recent segmental duplications in the working draft assembly of the brown Norway rat." Genome Res **14**(4): 493-506.

van de Sande, W. W., A. Fahal, et al. (2007). "Polymorphisms in genes involved in innate immunity predispose toward mycetoma susceptibility." J Immunol **179**(5): 3065-3074.

VassarStats. (2010). "Vassar Statistical Calculator." Retrieved 6th February, 2010, from http://faculty.vassar.edu/lowry/VassarStats.html.

Vaupel, P., F. Kallinowski, et al. (1989). "Blood flow, oxygen and nutrient supply, and metabolic microenvironment of human tumors: a review." Cancer Res **49**(23): 6449-6465.

Vaux, E. C., S. M. Wood, et al. (2001). "Selection of mutant CHO cells with constitutive activation of the HIF system and inactivation of the von Hippel-Lindau tumor suppressor." J Biol Chem **276**(47): 44323-44330.

Venables, J. P., L. Strain, et al. (2006). "Atypical haemolytic uraemic syndrome associated with a hybrid complement gene." PLoS Med **3**(10): e431.

Viault, F. (1891). Comptes Rendu Hebdomadaire des séances de l'Académie des Sciences **112**: 295-300.

Vik, D. P. and D. T. Fearon (1987). "Cellular distribution of complement receptor type 4 (CR4): expression on human platelets." J Immunol **138**(1): 254-258.

Vissers, M. C., S. P. Gunningham, et al. (2007). "Modulation of hypoxia-inducible factor-1 alpha in cultured primary cells by intracellular ascorbate." Free Radic Biol Med **42**(6): 765-772.

Vissers, M. C. and R. P. Wilkie (2007). "Ascorbate deficiency results in impaired neutrophil apoptosis and clearance and is associated with up-regulation of hypoxia-inducible factor 1alpha." J Leukoc Biol **81**(5): 1236-1244.

von Fodor, J. (1887). "Die Faehigkeit des Bluts Bacterien zu vernichten." Deutsch. Med. Wschr. **13**: 745.

Voskarides, K., L. Damianou, et al. (2007). "COL4A3/COL4A4 mutations producing focal segmental glomerulosclerosis and renal failure in thin basement membrane nephropathy." J Am Soc Nephrol **18**(11): 3004-3016.

Vyse, T. J., B. J. Morley, et al. (1996). "The molecular basis of hereditary complement factor I deficiency." J Clin Invest **97**(4): 925-933.

Vyse, T. J., P. J. Spath, et al. (1994). "Hereditary complement factor I deficiency." QJM **87**(7): 385-401.

Wallis, R., D. A. Mitchell, et al. (2010). "Paths reunited: Initiation of the classical and lectin pathways of complement activation." Immunobiology **215**(1): 1-11.

Walmsley, S. R., A. S. Cowburn, et al. (2006). "Neutrophils from patients with heterozygous germline mutations in the von Hippel Lindau protein (pVHL) display delayed apoptosis and enhanced bacterial phagocytosis." Blood **108**(9): 3176-3178.

Walport, M. J. (2001). "Complement. First of two parts." N Engl J Med **344**(14): 1058-1066.

Wang, G. L. and G. L. Semenza (1993). "Characterization of hypoxia-inducible factor 1 and regulation of DNA binding activity by hypoxia." J Biol Chem **268**(29): 21513-21518.

Wang, G. L. and G. L. Semenza (1993). "General involvement of hypoxia-inducible factor 1 in transcriptional response to hypoxia." Proc Natl Acad Sci U S A **90**(9): 4304-4308.

Wang, W., C. P. Winlove, et al. (2003). "Oxygen partial pressure in outer layers of skin of human finger nail folds." J Physiol **549**(Pt 3): 855-863.

Warnecke, C., Z. Zaborowska, et al. (2004). "Differentiating the functional role of hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha (EPAS-1) by the use of RNA interference: erythropoietin is a HIF-2alpha target gene in Hep3B and Kelly cells." FASEB J **18**(12): 1462-1464.

Webb, J. D., M. L. Coleman, et al. (2009). "Hypoxia, hypoxia-inducible factors (HIF), HIF hydroxylases and oxygen sensing." Cell Mol Life Sci **66**(22): 3539-3554.

Weeks, D. E., J. Ott, et al. (1990). "SLINK: a general simulation program for linkage analysis." Am J Hum Genet **47**: A204.

Weiler, J. M., M. R. Daha, et al. (1976). "Control of the amplification convertase of complement by the plasma protein beta1H." Proc Natl Acad Sci U S A **73**(9): 3268-3272.

Weis, J. J., T. F. Tedder, et al. (1984). "Identification of a 145,000 Mr membrane protein as the C3d receptor (CR2) of human B lymphocytes." Proc Natl Acad Sci U S A **81**(3): 881-885.

Wellcome Trust Case Control Consortium (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-678.

Wiesener, M. S., M. Seyfarth, et al. (2002). "Paraneoplastic erythrocytosis associated with an inactivating point mutation of the von Hippel-Lindau gene in a renal cell carcinoma." Blood **99**(10): 3562-3565.

Willam, C., P. H. Maxwell, et al. (2006). "HIF prolyl hydroxylases in the rat; organ distribution and changes in expression following hypoxia and coronary artery ligation." J Mol Cell Cardiol **41**(1): 68-77.

Williams, T. N. (2006). "Red blood cell defects and malaria." Mol Biochem Parasitol **149**(2): 121-127.

Williams, T. N., T. W. Mwangi, et al. (2005). "Negative epistasis between the malaria-protective effects of alpha+-thalassemia and the sickle cell trait." Nat Genet **37**(11): 1253-1257.

Williams, T. N., S. Wambua, et al. (2005). "Both heterozygous and homozygous alpha+ thalassemias protect against severe and fatal Plasmodium falciparum malaria on the coast of Kenya." Blood **106**(1): 368-371.

Wilschanski, M., Y. Schlesinger, et al. (2007). "Combination of Helicobacter pylori strain and tumor necrosis factor-alpha polymorphism of the host increases the risk of peptic ulcer disease in children." J Pediatr Gastroenterol Nutr **45**(2): 199-203.

Wilson, C. B. (1986). "Immunologic basis for increased susceptibility of the neonate to infection." J Pediatr **108**(1): 1-12.

Wilson, E. B. (1927). "Probable inference, the law of succession, and statistical inference." J Am Statistical Association **22**: 209-212.

Winearls, C. G. (1995). "Historical review on the use of recombinant human erythropoietin in chronic renal failure." Nephrol Dial Transplant **10 Suppl 2**: 3-9.

Winearls, C. G., D. O. Oliver, et al. (1986). "Effect of human erythropoietin derived from recombinant DNA on the anaemia of patients maintained by chronic haemodialysis." Lancet **2**(8517): 1175-1178.

Wojta, J., K. Huber, et al. (2003). "New aspects in thrombotic research: complement induced switch in mast cells from a profibrinolytic to a prothrombotic phenotype." Pathophysiol Haemost Thromb **33**(5-6): 438-441.

Wu, H., S. A. Boackle, et al. (2007). "Association of a common complement receptor 2 haplotype with increased risk of systemic lupus erythematosus." Proc Natl Acad Sci U S A **104**(10): 3961-3966.

Wykoff, C. C., C. W. Pugh, et al. (2000). "Identification of novel hypoxia dependent and independent target genes of the von Hippel-Lindau (VHL) tumour suppressor by mRNA differential expression profiling." Oncogene **19**(54): 6297-6305.

Yamashina, M., E. Ueda, et al. (1990). "Inherited complete deficiency of 20-kilodalton homologous restriction factor (CD59) as a cause of paroxysmal nocturnal hemoglobinuria." N Engl J Med **323**(17): 1184-1189.

Yamashita, T., O. Ohneda, et al. (2008). "The microenvironment for erythropoiesis is regulated by HIF-2alpha through VCAM-1 in endothelial cells." Blood **112**(4): 1482-1492.

Yau, S. C., M. Bobrow, et al. (1996). "Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis." J Med Genet **33**(7): 550-558.

Ye, S., S. Dhillon, et al. (2001). "An efficient procedure for genotyping single nucleotide polymorphisms." Nucleic Acids Res **29**(17): E88-88.

Yefenof, E., G. Klein, et al. (1976). "Surface markers on human B and T-lymphocytes. IX. Two-color immunofluorescence studies on the association between ebv receptors and complement receptors on the surface of lymphoid cell lines." Int J Cancer **17**(6): 693-700.

Yeo, E. J., Y. S. Cho, et al. (2008). "Contribution of HIF-1alpha or HIF-2alpha to erythropoietin expression: in vivo evidence based on chromatin immunoprecipitation." Ann Hematol **87**(1): 11-17.

Yu, A. Y., L. A. Shimoda, et al. (1999). "Impaired physiological responses to chronic hypoxia in mice partially deficient for hypoxia-inducible factor 1alpha." J Clin Invest **103**(5): 691-696.

Zhang, H., P. Gao, et al. (2007). "HIF-1 inhibits mitochondrial biogenesis and cellular respiration in VHL-deficient renal cell carcinoma by repression of C-MYC activity." Cancer Cell **11**(5): 407-420.

Zhou, J., T. Schmid, et al. (2004). "PI3K/Akt is required for heat shock proteins to protect hypoxia-inducible factor 1alpha from pVHL-independent degradation." J Biol Chem **279**(14): 13506-13513.

Zimmer, M., D. Doucette, et al. (2004). "Inhibition of hypoxia-inducible factor is sufficient for growth suppression of VHL-/- tumors." Mol Cancer Res **2**(2): 89-95.

Zipfel, P. F. and C. Skerka (1994). "Complement factor H and related proteins: an expanding family of complement-regulatory proteins?" Immunol Today **15**(3): 121-126.