A study of inherited kidney disease

Thomas MF Connor

University College London

Thesis submitted for the degree of Doctor of Philosophy 2013

Declaration

I, Thomas Connor, 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 many people who have helped me over the years. First, I would like to thank my supervisors, Patrick Maxwell and Margaret Ashcroft, for their on-going support, enthusiasm, and, above all, for giving me the chance to take this on. Particular thanks are also due to Danny Gale, Guy Neild, and Patrick Chinnery for their advice and mentorship.

I am very grateful to the many past and present members of the Maxwell and Ashcroft groups who have helped me along the way. I have benefited from the experience of a great many excellent post-docs: Sarah Harten, Deepa Shukla, Evon Poon, Oliver Staples, Katie Ryan, Luke Thomas, Afshan Ahmed. I am also very grateful to the continuing support of my fellow students: Nick Annear, Tom Briston – for doughnuts and sourdough.

I would like to thank Kerra Pearce for help with SNP genotyping, Konstantinos Voskarides, for generous access to Cypriot DNA, and Gavin Hudson for help with mitochondrial sequencing.

I am especially grateful to the many patients and family members who have contributed to the Genetic Investigation of Kidney Disease study, and without whom this research would not have been possible

Lastly to my family, to Tilly, Ben, and Meg, who keep me going, and my parents, Tim and Lou, who made me think this might be possible.

Abstract

Kidney disease is a common, expensive, and growing worldwide health problem. Genetic factors play an important role in the aetiology of kidney disease. Current research suggests that these genetic factors are predominantly rare variants with large phenotypic effects. In this thesis I have used a range of genetic techniques to identify rare variants in different families with kidney disease, and to study how they might cause disease.

The Turkish-Cypriot population of Northern Cyprus has a high incidence of renal disease, much of which is undiagnosed and may be inherited. I collected DNA from the entire population on renal replacement therapy and identified three individuals with the G871C mutation in *COL4A3*. I used conflicting homozygosity analysis to demonstrate a minimal shared haplotype, thus dating this mutation to 17 generations ago.

I used linkage analysis and whole genome sequencing in a large Greek-Cypriot kindred to identify 3 novel non-synonymous variants associated with kidney disease. Expression of these variants was examined in cultured primary urothelial cells from this family.

I have studied another large pedigree with maternal transmission of renal disease. Sequencing of the mitochondrial genome demonstrated the presence of a novel polymorphism in the heavy strand promoter region at homoplastic levels. Mitochondrial function in primary dermal fibroblasts demonstrated a significant reduction in baseline oxidative respiration with a compensatory increase in glycolysis.

Lastly, I have studied a novel compound heterozygous mutation in *VHL*. This variant showed abnormal degradation of HIF without activation of HIF target genes in patient-derived B-cells. It is possible that these cells are able to employ some kind of VHL-independent HIF regulatory mechanism.

These studies demonstrate, in differing ways, the challenges of linking phenotype to genotype. Understanding the pathological and therapeutic importance of genetic information will become increasingly important to our management of kidney disease in this post-genomic era.

Table of contents

Table of contents

Table of contents

List of figures

List of figures

List of figures

List of tables

List of tables

List of abbreviations

ACE2	Angiotensin-converting-enzyme 2
ACR	Albumin creatinine ratio
ADPKD	Autosomal dominant polycystic kidney disease
Akt	v-akt murine thymoma viral oncogene homolog 1
ALMS1	Alstoms syndrome 1
APOL1	Apolipoprotein 1
ARMS	Amplification Refractory Mutation System
ATP	Adenosine triphosphate
B2M	Beta-2-microglobulin
CAKUT	Congenital abnormalities of the kidney and urological tract
CCRCC	Clear cell renal cell carcinoma
CDH1	E-cadherin
CFHR5	Complement factor H-related 5
CH	Conflicting homozygosity
CKD	Chronic kidney disease
COL4A3	Collagen, type-IV, alpha-3
COL4A4	Collagen, type-IV, alpha-4
COL4A5	Collagen, type-IV, alpha-5
CTS3	Cystatin 3
CytC	Cytochrome C
DACH1	Dachshund homolog 1
DFO	Desferrioxamine
DHX32	DEAH (Asp-Glu-Ala-His) box polypeptide 32
DMOG	Dimethyloxalylglycine
DNB	DNA nanoballs
ECAR	Extracellular acidification rate
EM	Electron microscopy
EPO	Erythropoietin
ERA-EDTA	European Renal Association - European Dialysis and Transplant Association
ESRD	End-stage renal disease
FCCP	Carbonyl cyanide p-[trifluoromethoxy]-phenyl-hydrazone

List of abbreviations

FSGS	Focal segmental glomerulosclerosis
FXN	Frataxin
GDP	Gross domestic product
GFR	Glomerular filtration rate
GLUT1	Glucose transporter 1
GSK-3ß	Glycogen synthase kinase 3 beta
GWAS	Genome-wide association study
HBF1 ß	Hepatocyte nuclear factor 1 homeobox B
HIF	Hypoxia inducible factor
HRE	Hypoxia response elements
HSP	Heavy strand promoter
IBD	Identical by descent
IF	Immunofluorescence
INPP5E	Inositol polyphosphate-5-phosphatase
INPP5F	Inositol polyphosphate 5-phosphatase F
KRT7	Keratin 7
L1CAM	L1 cell adhesion molecule
LCL	Lymphoblastoid cell line
LD	Linkage disequilibrium
LOD	Logarithm of the odds
LSP	Light strand promoter
MAF	Minor allele frequency
MCKD1	Medullary cystic kidney disease 1
MDM2	p53 E3 ubiquitin protein ligase homolog
MEFV	Pyrin
MRCA	Most recent common ancestor
mtDNA	Mitochondrial DNA
MYH9	Non-muscle myosin heavy chain 9
NAC	N-acetyl-L-cysteine
NADH	Nicotinamide adenine dinucleotide phosphate
NCL	Neuronal ceroid lipofuscinoses
ND1	NADH dehydrogenase subunit 1
NPHS1	Nephrin
NPHS2	Podocin

List of abbreviations

OCR	Oxygen consumption rate
OCRL	Oculocerebrorenal syndrome of Lowe
ODD	Oxygen-dependent degradation
P4HA1	Prolyl-4 hydroxylase alpha-1
PAI1	Plasminogen activator inhibitor-1
PAS	Periodic acid Schiff
PCDH15	Protocadherin-15
PFKFB4	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4
PHD	Prolyl hydroxylase
PKD1	Polycystin-1
PKD2	Polycystin-2
PLA2R1	phospholipase A2 receptor 1
PLCE1	Phospholipase C epsilon 1
RACK1	Receptor of activated protein kinase C 1
RCC	Renal cell carcinoma
RNR1	Mitochondrially encoded 12S RNA
RNR2	Mitochondrially encoded 16S RNA
ROS	Reactive oxygen species
RRT	Renal replacement therapy
SHROOM3	Shroom family member 3
SNP	Single nucleotide polymorphism
TBMN	Thin basement membrane nephropathy
TCA	Tricarboxylic acid
TFAM	Mitochondrial transcription factor A
TSC2	Tuberous sclerosis 2
UMOD	Uromodulin
VDAC1	Voltage-dependent anion channel 1
VEGF	Vascular endothelial growth factor
VHL	von Hippel-Lindau
WGS	Whole genome sequencing

Chapter 1 Introduction

1.1 Overview

My thesis focuses on the genetics of kidney disease. I have studied the Turkish-Cypriot population of Northern Cyprus, which has a high incidence of familial renal disease. I have studied families in which many members have kidney disease that has progressed to end-stage renal disease (ESRD). I have identified a number of genetic polymorphisms that co-segregate with kidney disease in these families, and I demonstrate mechanistic work to show how these polymorphisms cause disease.

In the introduction I will explain some of the principles of renal genetics. Kidney disease is a common and increasing worldwide health problem with significant financial implications, and kidney cancer is among the most common adult malignancies. In both cases, genetic factors play an important role. Moreover, in both cases, the discovery of genes responsible for rare monogenetic syndromes has led to important developments in our understanding of the underlying pathophysiology. I will briefly discuss two illustrative cases; firstly autosomal dominant polycystic kidney disease and secondly von Hippel-Lindau disease.

Linkage analysis has been used for many years to map genetic loci through the study of related individuals. This process has been simplified by the use of high density DNA arrays that give highly accurate estimates of recombination and measures of identity by descent. More recently, next generation sequencing has enabled the rapid, and now cost-effective, assessment of all sequence variants in an individual. Two key problems remain, however: the quality of phenotypic information and the challenge of showing how a given sequence variant actually causes the observed kidney phenotype.

While it is clear that the aetiology of most complex diseases has a genetic component, the process of elucidating this has been slow and difficult. Linkage analysis has focused on rare, distinctive phenotypes, while the approach of association studies has mostly uncovered variants that confer only small increments in risk [1] (Figure 1.1). The revolution in genomic technology has resulted in an exponential increase in our ability to investigate genetic risk factors [2]. In this post-genomic era it is hoped, therefore, that the study of rare families, especially those occurring in high risk populations, will help to find the remaining heritability of kidney disease.

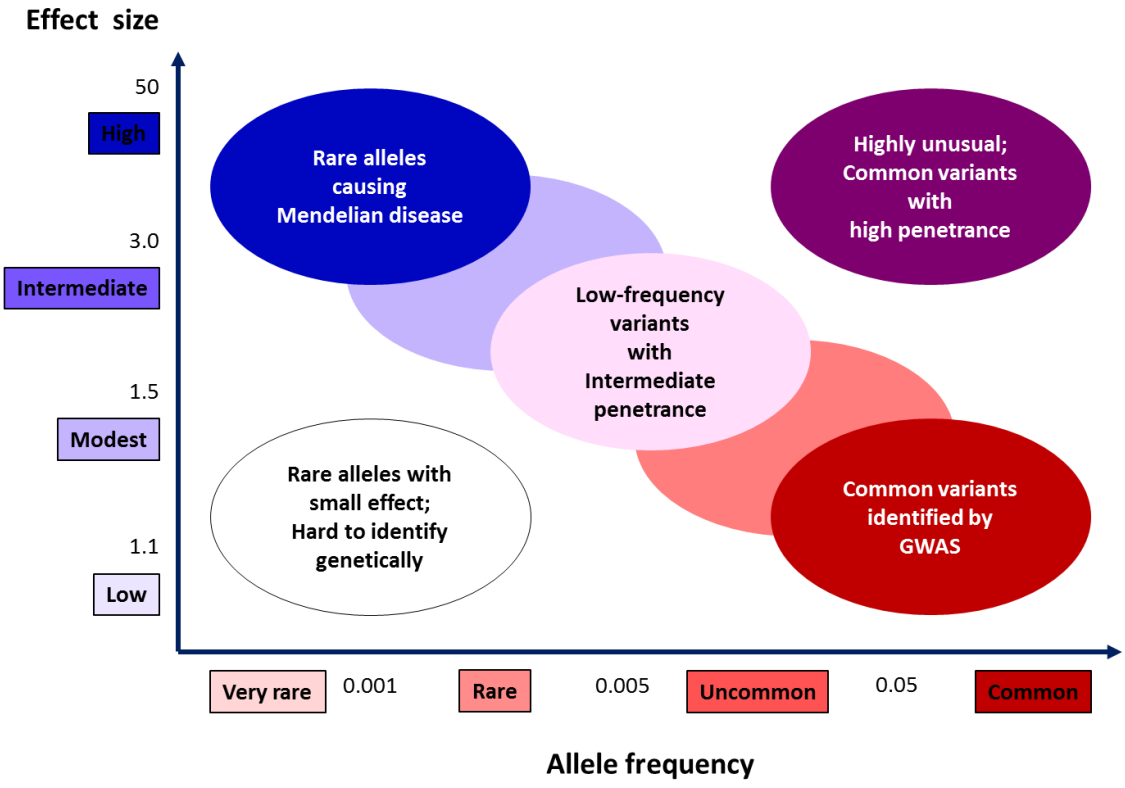


Figure 1.1 Identification of genetic variants by allele frequency and effect size

Linkage analysis has been used to identify highly penetrant alleles with large effect sizes (at the upper end of the figure), whereas genome-wide association studies (GWAS) have been powered to identify common alleles (at the right side of the figure). Adapted from [1].

1.2 Background to kidney disease

Chronic kidney disease (CKD) is a worldwide health problem, affecting an estimated 5-10% of adults in most populations that have been studied [3-5]. The prevalence is increasing, due to rising rates of diabetes mellitus, obesity, hypertension, and aging populations [6]. There is a strong inverse correlation between mortality and kidney function with more advanced kidney disease (GFR <45 ml/min) [7, 8], mainly due to increased cardiovascular risk [9]. Individuals with impaired kidney function are also at increased risk of progression to end-stage renal disease (ESRD) [10, 11].

Although only a small proportion of patients progress to ESRD, the cost of renal replacement therapy (RRT) makes this 0.1% of the population responsible for 1-2% of health-care spending in high-income countries [12]. Given that both the incidence of ESRD and cost of its treatment

are increasing, this has become an increasing public health challenge for high and middle income countries. ESRD is of course not a single disease process, but rather represents the final common path of a wide variety of pathophysiological processes.

Kidney disease encompasses a wide range of disorders, resulting from the complex interplay of many known and unknown environmental and genetic factors. The genetic contribution to kidney function and kidney disease is supported by a number of lines of evidence. First, there are a large number of monogenetic diseases of the kidney, such as autosomal dominant polycystic kidney disease (ADPKD) or Alport's disease, that are important aetiologies for ESRD. Second, there are a number of studies showing significant heritability of a number of commonly used markers of renal disease. Third, CKD shows familial aggregation and clustering by race, which probably reflects both shared environmental and genetic susceptibility.

The last four decades have seen the emergence of diabetes and hypertension as the dominant cause of ESRD [13-15], reflecting both improved survival of diabetic and hypertensive patients and more liberal access to RRT [16]. However, CKD progression in individuals with hypertension and /or diabetes is variable, pointing towards the importance of additional risk factors, including genetic risk factors. Moreover, familial clustering amongst incident dialysis patients is greatest for those where ESRD is caused by diabetes or hypertension [17].

Current treatment of kidney disease is designed to modify underlying risk factors, such as diabetes or hypertension, and reduce activation of the renin-angiotensin system. Whilst this has undoubtedly slowed the rising incidence of ESRD [14], residual proteinuria is common and there is wide variation in the response to treatment. In some cases, especially where there is active inflammation, specific treatments are successful. However in the great majority of cases, current therapeutic options do not address the underlying pathophysiology of renal disease. It is hoped that better understanding of the genetic contribution to kidney function and kidney disease will allow the development of novel therapeutic approaches that will delay or even prevent the progression to ESRD.

1.3 Kidney cancer background

In addition to studying the phenotype of progressive kidney failure, this thesis also examines the phenotype of kidney cancer. Kidney cancer is among the most common adult malignancies. It accounts for over 3% of all new cases of cancer diagnosed in men and around 2% of all cancers in women in the UK [18]. In the US, one in 75 people will develop renal cancer in their lifetime and approximately one-third will have metastatic disease at presentation. Kidney cancer has a notoriously poor response to chemotherapy and radiotherapy, but treatment has evolved significantly in the past 10 years. Key to these recent developments in therapy has been a revolution in our understanding of the molecular basis of renal tumour syndromes [19].

Much of our understanding of the molecular basis of renal cell carcinoma (RCC) has come from the study of a number of familial cancer syndromes. Inherited RCC is characterised by an earlier age at diagnosis than in sporadic cases, and it is often multi-centric or bilateral. Although only 2-3% of all cases of RCC are familial [20]; the identification of such cases is important to allow screening of the affected individuals [21]. The presence of characteristic tumours or extra-renal signs may point to a diagnosis of a specific RCC susceptibility syndrome.

Subtype	**Percentage of all renal cancers**	**Syndrome**	**Mutation**
Clear Cell	75-80%	Von Hippel-Lindau disease Familial Paraganglioma	*VHL* (loss of function) *SDHB* (loss of function)
Papillary Type I	10-15% (types 1 & 2)	Hereditary Papillary Renal Carcinoma	*MET* (activating)
Papillary Type II		Hereditary Leiomyomatosis and Renal Cancer	*FH* (loss of function)
Chromophobe +/- Oncocytic	5%	Birt Hogg Dubé	*FLCN* (loss of function)
Angiomyolipoma	<1%	Tuberous Sclerosis Complex	*TSC1* & *TSC2* (loss of function)
Collecting Duct	<1%	Lynch syndrome	Unknown

Table 1.1 Genetic causes of familial renal cancer

To date there are seven genes that have been implicated in familial renal carcinoma (Table 1.1). These genes were identified by positional cloning on the basis of characteristic histology and significant extra-renal signs; however they have been shown to be important in the much wider population of sporadic tumours. These genes are all implicated in two interrelated metabolic pathways, which mediate the cellular response to changes in oxygen tension and the availability of energy substrates. Interestingly, mutations of these genes results in unrestricted proliferation of renal epithelial cells, but not the progression to ESRD caused by loss of normal physiological function. The most common cause of inherited RCC is von Hippel–Lindau (VHL) disease, which is caused by mutations in the *VHL* tumour suppressor gene.

1.4 Evidence for genetic factors in kidney disease

Chronic kidney disease and the progression to end stage renal disease is not a single disease process. There are wide variations in incidence and prevalence both nationally and internationally, reflecting a wide range of contributory factors. Broadly speaking, phenotypic variation in a population arises from two sources: genotypic differences (nature) and environmental factors (nurture). Although CKD is itself a complex phenotype, there are a number of lines of evidence that genetic factors play a critical role in its aetiology.

A significant challenge for this research, however, is that we are compelled to study an intermediate phenotype [22]. In the absence of a detailed understanding of its complex pathophysiology, CKD must be disassembled into components that may be more readily measured. These components include such traits as albumin excretion, which reflects increased glomerular permeability, and serum creatinine concentration, which reflects glomerular filtration rate. Direct analysis of renal tissue would enable assessment of more specific phenotypes. While this is sometimes possible in small-scale family studies, it is impossible across the large numbers of mostly unaffected individuals needed for a population scale study. Moreover, these traits may themselves be more closely attributable to underlying genetic variation than the disease phenotype.

1.4.01 Monogenetic kidney disease

There are a significant number of rare kidney diseases that are caused by single gene defects. The spectrum of pathology here incorporates almost every single type of renal pathology, from developmental abnormalities to the haemolytic uraemic syndrome [23-25]. Many of these conditions have significant carrier rates in the general population. Thus, mutations in type 4 collagen, responsible for thin basement membrane nephropathy and Alport's syndrome may occur in up to 1% of the population [26], and autosomal dominant polycystic kidney disease is one of the most common monogenetic disorders, with an incidence of 1:400 – 1:1000 [27].

The genes involved in these monogenetic disorders encode proteins involved in all segments of the nephron. Their discovery in the last 30 years has often provided a starting point for the elucidation of the related disease pathogenesis [25]. This has led to the understanding of the critical role played by the glomerular podocyte in the maintenance of the glomerular filtration barrier, the role of numerous transporter and channel proteins in various types of tubulopathies, and the importance of the primary cilium in renal cystic disease.

These monogenetic disorders exhibit various patterns of inheritance and age of onset. Moreover, the demonstration of genetic and clinical heterogeneity has confirmed that a significant proportion of simple monogenetic disorders are far more complex than previously thought. This is especially true for developmental and cystic renal disorders. Renal developmental disorders under the acronym CAKUT (congenital abnormalities of the kidney and urological tract), such as those caused by mutations in *PAX2* and *HNF-1B*, exhibit a wide spectrum of abnormalities in different individuals carrying the same mutation [25]. Similarly, mutations in single ciliary genes are often associated with multiple cystic phenotypes. Conditions such as Bardet-Biedl syndrome are caused by mutations in multiple genes, and the severity of such conditions is associated both with the type and number of mutations, a concept called mutational load [28].

1.4.02 Heritability studies

Another argument for the importance of genetic factors in the development of renal disease comes from heritability studies. These studies attempt to quantify the relative contribution of environment and genetic factors, although just as there are several ways to define genetic influences, there are many measures of heritability (Figure 1.2). In the broadest sense heritability refers to the total genetic contribution, whereas in the narrowest sense, strict heritability refers to the additive genetic variance that drives the resemblance between relatives and can be quantified more easily [29]. Heritability estimates reflect the amount of variation in genotypic effects compared with variation in environmental effects, rather than absolute measures of the contribution of genetic and environmental factors to a phenotype. Estimates are therefore relative to population-specific factors, such as allele frequencies, measurable effects of gene variants, and the influence of environmental factors. Increasing genetic variation, as with outbreeding populations, or decreasing environmental variation, as with controlled experimental conditions, can increase heritability estimates. Nevertheless, in practice, heritability estimates of similar traits are often remarkably similar across populations [30].

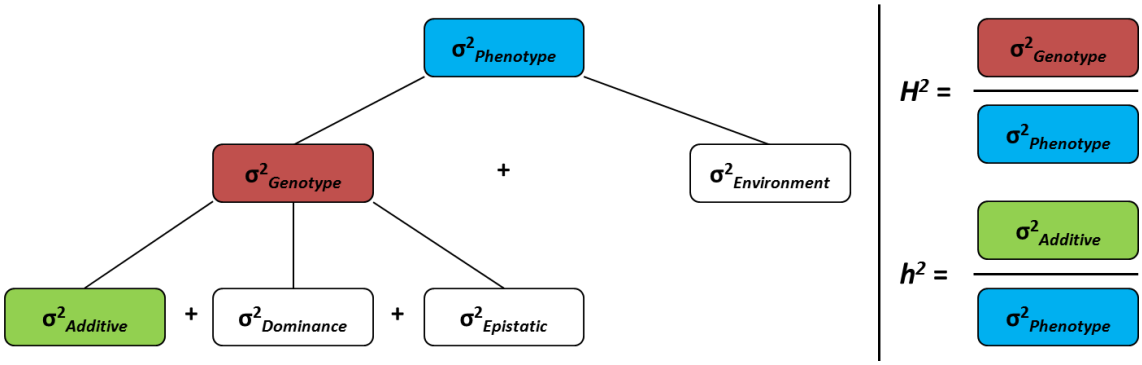


Figure 1.2 Estimation of heritability

Broad-sense heritability (H^2) refers to the total genetic contribution to a given phenotype. Strict, or narrow-sense heritability (h^2) specifically refers to additive genetic variance. Resemblance between relatives is mostly driven by additive genetic variance; therefore narrow sense heritability tells us how well a parent's phenotype will predict that of the offspring. By contrast, H^2 tells us what proportion of phenotypic variation is caused by the genotypes of the individuals in the population. Adapted from [29].

Heritability estimates for the most commonly used measure of kidney function, glomerular filtration rate (GFR), range from 0.33 – 0.82. [31, 32]. This indicates that 33-82% of the inter-individual variance in GFR estimates in these studies could be explained by additive genetic effects, in other words, narrow sense heritability. Heritability estimates for urinary albumin excretion, a hallmark of diabetic nephropathy, range from 0.3 to 0.44 in Finnish, New England, and south-eastern US families enriched for members with type 2 diabetes [33-35]. Glomerular filtration rate (GFR) was similarly heritable in families with diabetes and hypertension, with estimates ranging from 0.36 to 0.75 [32, 35, 36].

One problem with this kind of research is that kidney disease is not a tightly defined phenotype. There are many available measure of kidney function, from inulin clearance to estimated GFR and cystatin C measurements, which appear to exhibit differing degrees of heritability. Furthermore, GFR changes substantially over the human lifespan. This is a recurrent problem with population-based approaches to assessing the genetic risk of kidney disease, and will be discussed further in the context of genome-wide association studies.

Studies of heritability are often conducted in isolated, endogamous populations, such as the Zuni Indians of New Mexico [37], or the families of African descent in the Seychelles [32], where there disease phenotype under study occurs at significant levels. However, the estimates of heritability are lower in outbred populations, such as that of the Framingham Heart Study [31, 34]. The advantage of using isolated populations to estimate heritability is that they show a shared stable environment and reduced genetic complexity [38].

The population structure of a genetic isolate can be variable, ranging from the populations of Sardinia and Finland to tiny isolates reported in the South Tyrol and Val Borbera in Italy or in the Netherlands [38]. Isolated founder populations originating from recent common ancestors exhibit extended linkage disequilibrium relative to the general population. The Val Borbera population of northern Italy shows that the heritability of serum creatinine is less than, for example, total cholesterol or mean cell haemoglobin, however it is still in the region of 20-35% [38]. In addition, these researchers noted that individuals presenting with extreme values for these traits exhibited a higher kinship coefficient, a measure of inbreeding, compared with individuals in other percentile groups.

Inbred populations offer a similar theoretic advantage to geographical isolates for the study of complex disease [39]. Such populations show higher levels of homozygosity in recessive

alleles, which has been shown to be harmful in almost all species studied [40]. This occurs across a wide range of traits and suggests a large number of deleterious alleles located throughout the human genome. Most studies of the effects of inbreeding have focussed mainly on its impact on reproduction, childhood mortality and rare monogenetic disorders [41]. However, natural selection would predict that the harmful effects of inbreeding would increase with age, and most genetic variants that have been identified causing complex disease in humans are partially recessive. Examination of genetic isolates on three islands in the eastern Adriatic revealed that 23-48% of complex disease could be attributed to recent inbreeding [39]. Whilst these studies did not directly measure renal function, studies in the endogamous Zuni Indians showed significant heritability of many phenotypes related to kidney disease.

1.4.03 Familial clustering

Familial aggregation studies show that ESRD, and earlier stages of CKD, clusters in families [17, 42]. Familial clustering has been reported in diabetic, hypertensive, systemic lupus erythematosis, and human immunodeficiency virus-associated nephropathies [43]. The presence of a family history of ESRD is a significant risk factor for the subsequent development of nephropathy, and the proportion of patients of RRT with a family history of ESRD is highest in young adults, non-Caucasians, and those where ESRD is caused by diabetes or hypertension.

It was many years before the risk of diabetic nephropathy and progression to diabetic ESRD was ascribed to anything other than the simple duration of exposure to hyperglycaemia [44]. However it is now clear that select individuals with diabetes are at differential risk for developing nephropathy, based on the familial aggregation of kidney disease [45]. This aggregation of diabetic nephropathy is independent of family size, the number of relatives affected with diabetes and hypertension, socioeconomic status, and inadequate access to health care, strongly suggesting the existence of susceptibility genes [42]. Analysis of nearly 26,000 incident US dialysis patients, excluding those with known genetic disorders, showed that nearly 32% of black women, 27% of black men reported having close relatives with ESRD, compared with 15% and 12% of European Americans, respectively [17]. Segregation analyses in multiplex families supports the importance of major genetic effects [45]. Interestingly,

familial aggregation of specific renal histological diagnoses is also observed in siblings with diabetes, independent of diabetes duration and disease control [46].

Hypertensive renal disease presents a similar picture to diabetic nephropathy; in that only a minority of hypertensive patients go on to develop kidney disease. Here again there is evidence of familial clustering and racial disparities in familial risk. Thus 19% of ESRD patients with hypertension in the south-eastern US had a first or second-degree relative with ESRD, and this risk was greater in African Americans than European Americans [45]. This risk is preserved, even when family history of hypertension and other potential risk-aggravating factors are controlled for [42]. Lastly, these population studies are supported by a number of animal models showing that the risk of renal damage due to genetic factors is independent from the presence of hypertension [22].

1.5 Identifying the genetic basis of kidney disease

Kidney disease is therefore a mixture of single gene defects with high phenotypic impact, and a combination of multiple genetic and environmental factors, with harmful environmental exposures acting on genetically susceptible individuals [47]. Our ability to uncover the genetic basis of kidney disease has been revolutionised by the recent explosion in genomic technology. There are two fundamental approaches to the identification of disease genes, linkage and association analysis. Initial progress was made with linkage analysis, studying highly penetrant monogenetic disorders in large pedigrees. Early association studies looked, without great success, at a series of candidate genes. Subsequently, unbiased, genome-wide association studies were performed to look for common variants affecting the incidence of common diseases, such as CKD. More recently, next generation sequencing is now becoming widespread, and holds out the promises and challenges of truly personalised medicine for the 21st century.

1.5.01 Linkage analysis

Linkage analysis can be used to identify regions of the genome that contain genes that predispose to disease [48]. This analysis uses pedigree information to see which genetic markers are co-inherited with a given clinical phenotype. Two genetic loci are linked if they are transmitted together from parent to offspring more often than expected by chance (Figure 1.3). In other words, two loci are linked if, during meiosis, recombination occurs between them with a probability of less than 50% [49]. The probability of recombination is therefore a function of chromosomal distance between two loci, measured in centimorgans (cM). Linkage analysis is a measure of mathematical probability, rather than being driven by a previous biologically driven hypothesis, and is therefore often the first stage in the genetic investigation of a trait.

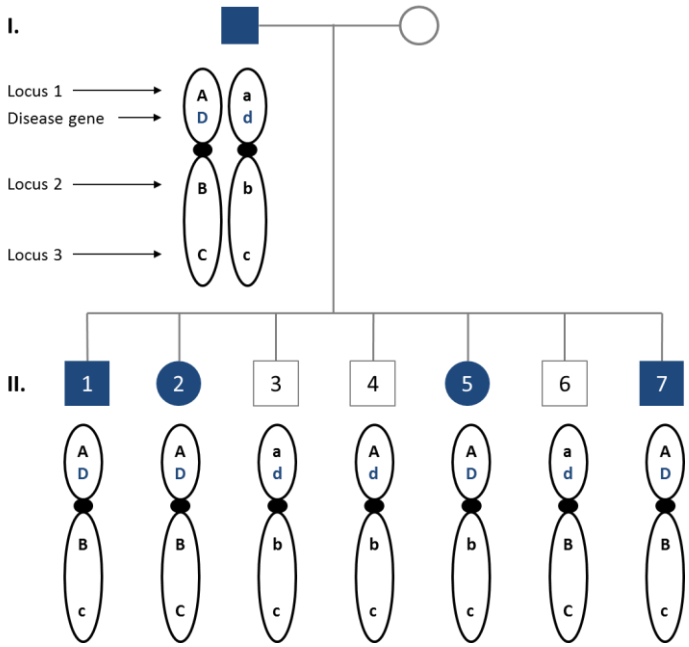


Figure 1.3 Co-inheritance of a disease gene (D) with three marker loci (A, B, and C)

Only the father's contribution to the genotype is shown. All four affected offspring have inherited the A allele along with D at locus 1. However II-1, II-5 and II-7 are recombinant for marker locus 3, and II-5 is also recombinant for marker locus 2, indicating that meiotic crossover has occurred between the two loci (adapted from [50]).

Parametric linkage analysis is the analysis of the co-segregation of genetic loci in pedigrees. It is possible to determine the position of genetic markers relative to each other and a disease locus by studying their segregation through a pedigree. The aim of such analysis is to determine the recombination fraction (θ), which is measured as an odds ratio of the two probability values (linked/unlinked). Linkage is usually reported as a $\log_{10}$, or logarithm of the odds (LOD) score. This score was first proposed by Morton in 1955 [51], and is a function of the recombination fraction (θ) or chromosomal position measured in cM. This means that the LOD score varies depending on the value of θ. Large positive values are evidence of linkage, while large negative values are evidence against.

Early studies used a variety of genetic markers, commonly highly variable microsatellite repeats, and might have included two-point comparisons between marker and disease locus. It is now common practice to use a very set large of markers, whose position is well mapped, spread evenly throughout the genome. Using multiple markers entails multiple estimates of θ, with the result that localisation is more accurate but the threshold of significance needs be amended. By convention, a LOD score of 3 was regarded as significant evidence of linkage, equivalent to a $p = 0.001$ [52]. Likewise, a score of -2 or less equated to a definite rejection of linkage. More recent work, taking into account the dense single nucleotide polymorphism (SNP) markers used, suggests that a threshold of 3.3 should be used, equivalent to a genome-wide significance of 0.05 [53].

Parametric analysis assumes that the genetic model for the disease of interest is known. This model, which must be specified at the start of analysis, includes the frequency of the disease allele, mode of inheritance, marker allele frequencies, and a full marker map for each chromosome. These model parameter estimates are ideally taken from population-based studies of the disease. Importantly, where these assumptions are incorrect, parametric linkage analysis can be severely compromised. Many genetic conditions show variable penetrance or genetic heterogeneity with different modes of inheritance. It is possible to overcome these obstacles, as was seen with the localisation of the breast cancer genes, *BRCA1* and *BRCA2* [54, 55].

An alternative approach is to use non-parametric analysis where there is no clear mode of inheritance [48]. This can be useful for multifactorial diseases, where several genes may contribute to disease risk. These kinds of analysis are based on the assumption that there is

excess sharing of haplotypes that are identical by descent (IBD) between affected relatives in the region of a disease causing gene, irrespective of the mode of inheritance. Various methods have been developed using sibling pairs or other groups of relatives, using computational algorithms to cope with the complexity of large pedigrees and large number of markers.

The choice of phenotypic trait is a significant issue with linkage analysis [48]. In simple Mendelian traits it can be straightforward to identify affected and unaffected individuals, and even in a disease such as cancer once the diagnosis is based on pathological findings. However there are a number of conditions such as obesity or psychiatric illness where definitive diagnosis may depend on a particular threshold or constellation of distinct symptoms. Chronic kidney disease a good example of this as it is often defined by the elevation of selected blood or urinary markers above a certain threshold. There may be a good reason for simplifying a quantitative trait into a binary one for the purpose of analysis; however the use of an inappropriate threshold can result in a loss of power [56]. The absence of a clear definition of phenotype will lead to uncertainty about the classification of affected and unaffected individuals. Moreover, in parametric linkage analysis, misclassification of individuals can dramatically reduce the size of calculated LOD scores.

Finally, while linkage analysis has been used with great success for over half a century, it is only the first step toward identifying an exact disease locus. In Mendelian traits, crossover events often narrow the locus sufficiently to define a small region of interest. However, such regions may equally contain hundreds of genes, many of which are biologically plausible candidates. The process of locus identification then required painstaking Sanger sequencing of candidate genes within the linked interval. The inclusion of additional affected individuals or more families with the same phenotype could improve this process, as Morton had originally envisaged [51], by increasing the LOD score and by introducing new recombinations to narrow the linked interval. The availability of next generation sequencing has further revolutionised this process by providing a direct route to the identification of likely causative variants.

1.5.02 Association studies

The alternative to looking for genetic factors within families is to conduct this analysis at the level of populations by means of association studies. The first studies on disease association were published in the 1950's and 1960's looking at the association of conditions such as stomach cancer and pernicious anaemia with ABO blood types [57]. These techniques were first applied to the field of kidney disease to investigate the effect of variants in a number of candidate genes, such as angiotensin-converting enzyme and angiotensinogen [58]. However it was not until technical developments in assaying single nucleotide polymorphisms (SNPs) in the decade after 1997 that it became possible to conduct truly genome-wide, hypothesis-free analyses.

In order to understand the basis of association studies, it is important to describe the nature of human genetic variation. There are many ways in which the DNA molecule can vary between individuals. SNPs are the commonest class of genetic variation among individuals, with nearly 200 million currently reported; however there are many others including deletions, insertions, inversions, and copy number variants of varying size [59]. Genetic variants are typically described as either common or rare, depending on whether the minor allele frequency (MAF) in the human population is greater than 1% or not. The vast majority of variants are predicted to have no effect on phenotypic variation. Thus they are evolutionarily ancient and have achieved their frequency in the human population by chance alone.

SNPs and other neutral variants located in the same genomic interval are often correlated with one another in what is termed linkage disequilibrium (LD) (Figure 1.4). Studies of the human population, such as the international HapMap project, have shown that these groups of correlated SNPS can be broken down into just ~550,000 LD blocks for individuals of European or Asian ancestry and ~1,100,000 block for those of African ancestry. By genotyping the DNA sample of an individual with a 'tagging' SNP from each LD bin, it is possible to estimate over 80% of SNPs present at a frequency above 5% in the genome [59].

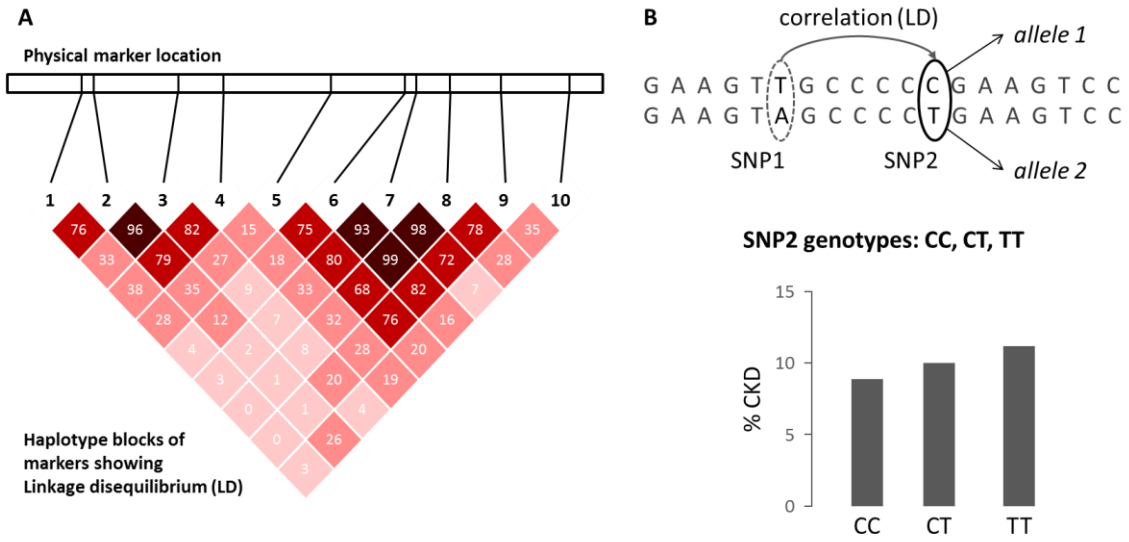


Figure 1.4 Linkage disequilbrium

Linkage disequilibrium (LD) is the non-random association of alleles at two or more markers, resulting in the occurrence of some combinations of alleles in a population more or less often than would be expected from a random formation of haplotypes based on their frequencies alone. (A) Red squared show high pairwise LD, thus markers 2-4 and 5-9 are in LD with each other, resulting in the formation of haplotype blocks. (B) Genome-wide association studies use single nucleotide polymorphism (SNP) markers to assess the association between genotypic variation and phenotypic variation. If SNPs 1 and 2 show complete LD, it is sufficient to only genotype alleles at SNP2 to assess the risk of disease caused by SNP1 (adapted from [47]).

Since the first GWAS using a high-throughput SNP genotyping array was published in 2005, a number of studies have been conducted looking at phenotypic definitions of CKD including an eGFR of <60 ml/min/1.73m^2 using the four variable Modification of Diet in Renal Disease Study equation, as well as GFR based on cystatin C and urinary albumin-to-creatinine ratio [47, 58]. These studies, employing pooled consortia with up to 67,000 participants, have revealed and replicated nearly 30 loci associated with population-based CKD, diabetic nephropathy, eGFR, and albuminuria [58]. These loci are in addition to those, such as *cystatin3* (*CST3)*, which were detected and have a clear relationship with the phenotype in question (in this case, production of cystatin C). They do, however, include many such as *shroom family member 3 (SHROOM3)* and *dachshund homolog 1 (DACH1)* which are expressed in the kidney and which appear to be important for normal development and function. Importantly, these loci also include a number, such as *uromodulin* (*UMOD)* and *alstrom syndrome 1 (ALMS1)* which have been implicated in rare Mendelian diseases of the kidney. This illustrates a key theme of findings from GWAS studies: common variants of genes responsible for severe, monogenetic disorders may be associated with milder presentations in the general population.

There are, however, a number of problems with the association study approach. One impetus behind the development of genome-wide association studies (GWAS) was the 'common disease-common variant' hypothesis, which proposes that common complex traits are largely due to common variants with small to modest effect sizes [60]. Because of the constraints imposed by multiple statistical tests and current sample sizes, GWAS are only able to investigate variants with a MAF of >1-5% [61]. Moreover, for the most part, GWAS can establish that a tag SNP from an LD block is statistically associated with a trait, but they cannot identify the precise variant(s) in the block that have the causal role in contributing to variation in the trait.

A good example of this limitations is given by a recent study using mapping by admixture linkage disequilibrium, which showed that the E-1 haplotype of nonmuscle myosin heavy chain 9 (*MYH9*) on 22q11-13 was a powerful risk factor for ESRD, as well as increasing the risk of hypertensive nephrosclerosis and focal segmental glomerulosclerosis [62, 63]. Variants in the adjacent apolipoprotein L1 (*APOL1*) gene were found to have an even stronger statistical association with ESRD than those of *MYH9*, and a plausible explanation for their evolutionary selection due to increased resistance to African sleeping sickness [64]. However, analysis of another MYH9 variant in Caucasians indicates than more than one renal risk variant might be operational in this LD block [58].

Another issue with association studies relates again to the issue of how the phenotype is defined. The first high-throughput GWAS in 2005 looked at age-related macular degeneration, and reported significant association with just 96 cases and 50 controls [65]. Since that time, the trend has been to create larger and larger study populations with the aim of increasing statistical power to detect small effects of variants in a large number of genes. However, it may well be that studies performed on a narrowly defined disease entity with a high genetic component of risk attributable to a single or limited number of genes may have more success; notably, they can succeed with a small study population along the lines of the study by Klein *et al* [65]. A recent association study of idiopathic membranous nephropathy provides a good example of this. Researchers took a cohort of just 556 patients with this condition (representing a significant proportion of all cases in the three countries that contributed to the study) and showed a highly significant association with variants in HLA-DQA1 and *phospholipase A2 receptor 1 (PLA2R1*) [66].

GWAS studies have made significant contributions to our understanding of the genetic basis of a number of aspects of kidney disease, and yet so far less than 2% of the estimated heritability has been explained [67]. This is a recurring theme across GWAS studies in a number of conditions, and has led to the concept of 'missing heritability' [1]. The remainder may be missing for a variety of reasons, both related to the size and stratification of the sample population, and to genetic factors such as epistatic interactions or allelic frequency that will be discussed subsequently.

1.5.03 Next generation sequencing

In 2007, the first complete genome sequence of an individual, J. Craig Venter, was published and shortly followed by that of a second individual, James D. Watson [68, 69]. These were significant advances on the reference versions of the human genome published in 2001, as these were both derived from numerous donors and represented the human genome as a haploid sequence without annotated genetic variation [59]. Since 2005 next generation sequencing platforms have become widely available, reducing the cost of DNA sequencing by four orders of magnitude relative to Sanger sequencing [70]. By coupling targeted capture and massively parallel DNA sequencing it has become possible to determine nearly all the coding variation present in an individual human genome, a process termed exome sequencing. This technique has become a powerful new approach for identifying the genes underlying Mendelian disorders. Moreover it offers the possibility to identify those rare variants which are thought to underlie some of the missing heritability.

Next generation sequencing is provided by a rapidly evolving set of technologies, however all platforms involve massively parallel sequencing of clonally amplified or single DNA molecules that are spatially separated in a flow cell [71]. This is a paradigm shift from that of Sanger sequencing, and the hundreds of megabases to gigabases of nucleotide sequence generated required an equal revolution in the computational pipelines used to assemble meaningful data. Targeted sequencing approaches have the general advantage of increased sequence coverage of regions of interest, although the different strategies of targeting capture introduce another layer of complexity and potential bias to the process.

The most immediate application of next generation sequencing is the discovery of the genetic defects underlying monogenetic inherited disorders. The number of rare monogenetic diseases is estimated to be over 5500, and yet for nearly 2000 of these conditions the underlying genes remain unknown (http://omim.org/statistics/entry). In addition, an increasing proportion of common disorders, such as schizophrenia or autism, that were previously thought to be due to complex multifactorial inheritance, are now thought to reflect a heterogeneous collection of rare monogenetic disorders [72]. Perhaps the most likely reason for missing heritability in GWAS studies described in the previous section is that many diseases are not a single disease caused by common variants with very low penetrance, but are actually a collection of different diseases, many of which have substantial contributions from rare variants with moderate or substantial phenotypic effects.

There are a number of strategies for finding disease-causing alleles, most of which rely on three discrete filtering steps which enable the reduction of nearly 2 million variants per individual to a minimum number of high-priority candidates. The first filtering step is used to eliminate candidate genes by assuming that any causative allele will not occur in a control population. This control population is often a public database, such as dbSNP or the 1000 Genomes Project, and represents a powerful approach because only around 2% of variants identified in an individual will be novel [73]. In addition to public databases, novelty can be assessed by filtering the variant against those found in a set of unaffected controls. This last is of particular relevance because most publically-available data was European in origin at the start, and might therefore not be so applicable to non-European or isolated populations.

This filtering step rests on the assumption that the control populations contain no alleles from individuals with the phenotype in question. However it is likely that reference databases, such as dbSNP, contain a small, but appreciable, number of pathogenic mutations [73]. This is due in part to incomplete phenotyping of the individuals whose variants are deposited in dbSNP, and in part to the second problem of allelic frequency in the population under study. As more variation data becomes publically available it is increasingly likely that pathogenic mutations that are present in the general population will be included, in direct proportion to the minor allelic frequency. This is especially relevant for recessive disorders, in which the carrier phenotype would not exclude an individual from a control population, but is equally true for dominant disorders with incomplete penetrance or advanced age of onset. With this caveat in

mind, it is better to filter out alleles above a MAF threshold of 0.1-1%, in conjunction with a larger sample size or in conjunction with pedigree-based approaches [73].

The second filtering process is the strategy by which the variants of a single individual can be cross-referenced with others to determine potentially pathological ones. The first successful use of next generation sequencing to identify a new Mendelian disorder in 2009 used comparison across four unrelated individuals with the same dramatic phenotype [74]. The number of individuals required is determined by what is known about the mode of inheritance and the possibility of genetic heterogeneity. Filtering is more efficient for recessive disorders because the genome of any one individual has around 50-fold fewer genes with two, rather than one, novel protein-altering alleles per gene [73].

Where possible, the use of pedigree information can significantly reduce the number of possible shared variants. The probability of identity by descent given identity by state is high for very rare alleles, even among distantly related individuals. It is therefore best to compare the most distantly related affected individuals within a pedigree, because each additional meiosis should on average reduce the shared genomic space by 50%. Where mapping data is available, this process can be refined to utilise those individuals whose shared haplotype is the smallest. Similarly, in recessive disorders arising in consanguineous pedigrees, the individual with least homozygosity should be targeted. An elaboration of this strategy is to look at parent-child trios for the presence of *de-novo* mutations. This approach is particularly applicable for disorders in which most cases are sporadic, and which are suspected to be dominantly inherited or in which substantial locus heterogeneity is expected.

Another challenge for next generation sequencing will be to identify rare variants that underlie complex traits. In the absence of pedigree data there have been two main approaches. The first has been to target those regions that have been identified by association studies. Based on the assumption that rare variants which influence a trait are in linkage disequilibrium with common variants that influence risk, this allows capture nearly all of the rare and very rare (MAF < 0.1%) alleles that are present in a sample that may be responsible for the association (Figure 1.4). The second approach is the extreme phenotype study design, in which individuals at either end of a phenotype distribution are selected for sequencing [75]. It is probable that it will eventually become standard to sequence all individuals in a disease cohort, however it

remains a major challenge that even with very large sample sizes the power to detect an association with a single rare variant is low (see discussion on GWAS).

The third filtering step is the stratification of candidates on the basis of their predicted impact or deleteriousness. For example, greater weight can be given to frame shifts, stop codons, and disruptions of canonical splice sites than to missense variants. Candidates can be stratified by existing biological or functional information about a gene, such as its role in a biological pathway or interactions with genes known to cause a similar phenotype. Where this information is not yet known, bioinformatics algorithms can be used to predict pathogenicity based on the observation that regions of genes and genomes in which mutations are deleterious tend to show high sequence conversion as a result of purifying selection [76]. It is possible to model conservation at both nucleotide and amino acid level, resulting in enrichment of variants which are likely to have a phenotypic impact.

Despite the increasingly sophisticated pipeline for the identification of causal variants using next generation sequencing, this approach can fail. Failures are usually either technical or analytical [73]. Technical failures arise because of problems with sequence coverage, base calling, or sequence alignment. Technical failures are likely to diminish rapidly over the coming years, and for now one strategy to minimise their impact is to utilise a control set of genomes that have been generated using the same sequencing technology. The use of targeted capture to generate exome data can be an additional source of error, given the need to define the target set for initial enrichment. Analytical failure can arise from any of the filtering steps discussed above. The major limitation here is genetic heterogeneity in which the same phenotype is produced by a number of different alleles. From an analytic perspective false negatives, in which the causative alleles are present in the control data set or exhibit reduced penetrance, result in a reduced signal-to-noise ratio that is indistinguishable from genetic heterogeneity.

In conclusion, next generation sequencing technology offers the opportunity to replace both linkage analysis and association studies in the near future. So far there has been most progress using exomic data to uncover the molecular basis of Mendelian conditions. The field will probably gradually move to whole genome sequencing as the cost of sequencing continues to fall and as our ability to make sense of non-coding variation develops [2, 73]. There remain significant technical, statistical and bioinformatic obstacles to the widespread use of this

technology in clinical diagnostics. However, even as these technical hurdles are being removed, it becomes even more critical to characterise accurately the phenotypes that are the raw material for this analysis. Lastly, it should be mentioned that the use of next generation sequencing for disease gene discovery involves a number of ethical issues, in particular the issue of consenting for sequencing research and the return of information, including incidental findings, to individuals and the wider scientific community.

1.6 Mechanisms of genetic kidney disease

It is clear therefore that a significant proportion of chronic kidney disease is due to genetic factors. Single gene disorders account for nearly 15% of individuals with ESRD in national registries [14, 15, 77], and have provided remarkable insights into the molecular pathology of renal pathophysiology. Genome-wide association studies have transformed our understanding of the genetic architecture of complex disease, but have singularly failed to explain its heritability by means of the common variant model [1, 47]. This is a direct result of the evolution of the human genome. The most likely explanation is that complex disease is caused by many rare alleles, operating below the threshold for a population wide effect.

The oldest human alleles originated in Africa, in parallel with the evolution of our species. These ancient polymorphisms are shared by all populations and account for approximately 90% of human variation [78]. Most of these variants have no functional significance and persist by chance alone in the absence of selection pressure. However, new mutations continually arise at a rate of approximately 175 per diploid human genome per generation [79]. The exponential growth of the human population has therefore resulted in a vast number of new alleles, which have generated an immense degree of genetic variation. These new alleles are rare because of their recent origin, and may be confined to just one person or family. Moreover they may not yet have been removed from the population by natural selection.

With such allelic diversity it is perhaps not surprising that genetic factors exhibit a remarkable degree of molecular heterogeneity. Each gene can harbour many hundreds or even thousands of rare, pathogenic variants in unrelated affected individuals. Moreover, the same mutation causing loss of function at a particular genetic locus, may lead to different phenotypes in

different individuals. Conversely, mutations in different genes in the same or related pathways may result in the same phenotype. These mechanisms are best illustrated by discussion of two illustrative cases; firstly autosomal dominant polycystic kidney disease and secondly von Hippel-Lindau disease.

1.6.01 Autosomal dominant polycystic kidney disease

Autosomal dominant polycystic kidney disease (ADPKD; Mendelian inheritance in man (MIM) 173900, 263200) is one of the most common of all monogenetic disorders, with an incidence of 1 in 400 to 1 in 1000 [27]. ADPKD is the single most common genetic cause of ESRD in the adult population, accounting for 6.5% of incident cases of ESRD in the UK. ADPKD occurs worldwide with similar incidence in different populations. It is a genetically heterogeneous condition with two genes (*PKD1* and *PKD2*) identified by positional cloning in 1994 and 1996 [80, 81].

Polycystin-1 (*PKD1)* is localised to the short arm of chromosome 16 immediately tail to tail with the tuberous sclerosis gene, *TSC2* [82]. This part of chromosome 16 exhibits complex duplication, and it was subsequently found to contain six non-coding homologues to *PKD1* (pseudo-genes, *PKD1 P1* through *P6*). These pseudo-genes share 97.7% identity with exons 1-33 of *PKD1*, and have made the process of molecular diagnostics considerably more difficult. *PKD2* was identified by homology between the predicted PKD1 and PKD2 proteins, highlighting it from others in the candidate region [81].

Mutations in *PKD1* account for ~85% of cases in clinically identified populations, and result in a more severe clinical phenotype. The average age of onset of ESRD is 20 years younger in patients with *PKD1* mutations compared to those with polycystin-2 (*PKD2)* mutations, however there is a large overlap and it is not possible clinically to differentiate between the two genes. ADPKD exhibits a high level of allelic heterogeneity, with a total of 619 pathogenic *PKD1* and 129 pathogenic *PKD2* mutations reported to date (ADPKD database [PKDB], http://pkdb.mayo.edu). These mutations are spread across the entire sequence of these genes, and the majority are private to a single pedigree. The majority of mutations are truncations, with an additional 25% missense mutations. The type of mutation does not seem to be strongly associated with the phenotype [83, 84].

The PKD1 protein is a large (4303 amino acid) integral membrane protein with 11 transmembrane domains and an extracellular region consisting of a variety of domains, which may be important for protein-protein interaction. The PKD2 protein is smaller and functions as a non-selective cation channel that transports calcium. Polycystin-1 and -2 interact to form a complex, whose major role is to regulate levels of intracellular calcium, located at the primary cilium [27]. Primary cilia are membrane-enclosed hair-like organelles that occur on the apical surface of renal tubular cells and on cells in many other organs. The primary cilium appears critical for the development of ADPKD as there are now over 50 genes that have been associated with ciliary disease and a cystic or dysplastic kidney phenotype [28].

Renal ciliopathies, including ADPKD, are often accompanied by abnormalities in other organs [28]. This is not surprising given that primary cilia occur almost ubiquitously throughout the human body. The pleiotropic phenotypes in these syndromic disorders can be linked to defects in ciliary function and provide an insight into the signalling pathways that require functional cilia to operate [27]. Many of these genes give rise to distinct, if overlapping, clinical characteristics, which is supported by the discovery of discrete protein sub-complexes within the cilium. However individuals from the same family can exhibit widely divergent clinical phenotypes, showing that the nature of the individual mutation is less important than other genetic and environmental factors.

The renal ciliopathies provide a good example of the concept of 'mutational load', showing how monogenetic disease may be influenced by other genes in the same pathway. Oligogenic inheritance, involving mutations in multiple genes, has been reported for most ciliopathies, after the first description in Bardet-Biedl syndrome [85, 86]. Emerging data suggest that ciliopathy-associated genes have the capacity to contribute pathogenic alleles to multiple ciliopathies, and there is evidence to suggest that the final phenotype is an overall measure of the disturbance of ciliary function[87]. In support of this, it has recently been shown that two hypomorphic *PKD1* or *PKD2* alleles may result in a more severe phenotype like autosomal recessive polycystic kidney disease *in utero* [88-90]. Moreover, the combination of *PKD1* mutations with other genes, such as *TSC2* and *HNF1β*, also results in a more severe manifestation [91].

The exact mechanism by which mutations in *PKD1* and *PKD2* cause ADPKD remains unknown. The greater severity of *PKD1* appears to be due to the development of more cysts at an early

age, not to faster growth [27]. Studies of conditional knockouts in mice indicate a developmental switch around postnatal day 13, after which time inactivation of *Pkd1* results in a much milder disease [92, 93]. In humans, the cells lining cysts show loss of heterozygosity in *PKD1* and *PKD2*, indicating a somatic second hit [94, 95]. Loss of polycystin-mediated calcium signalling affects a number of intracellular signalling pathways that may contribute to cyst proliferation. The number and developmental timing of second hits that accumulate in an individual may thus play an important role in the variable severity in pedigrees. Moreover, it is intriguing that this unrestricted proliferation is not associated with tumourigenisis, as will be seen with the case of VHL disease.

From a diagnostic perspective, ADPKD remains challenging, and molecular testing for ADPKD is not recommended in patients in whom a firm positive or negative diagnosis can be made by imaging alone [96]. Even after testing, 9% of individuals exhibit no mutation in either *PKD1* or *PKD2*. These patients tend to have milder disease and a negative family history, suggesting that they may have more subtle mutations, such as deep intronic changes that are missed by the exon-based screening approach [97]. Mosaicism has been described in two ADPKD families, and may be more prevalent in *de novo* cases. Some *PKD1* alleles may show incomplete penetrance [98], and there are several ADPKD families that are apparently not linked to *PKD1* or *PKD2*. Given these limitations, it is to be hoped that the application of next generation sequencing to molecular diagnostics through the combination of targeted re-sequencing and whole genome sequencing may finally resolve these issues [99].

In conclusion, ADPKD is associated with a low level of genetic heterogeneity, with two genes showing a high degree of penetrance and accounting for the vast majority of cases. However, *PKD1* especially exhibits extreme allelic heterogeneity, with many hundreds of mutations distributed across a large and highly repetitive gene. At a population level these equate to many rare variants conferring a high risk of ESRD. Investigation of the overlap with other ciliopathies has revolutionised our understanding of that structure, and suggested possible therapies. These ciliopathies also provide a clear example of how variants affecting multiple genes within a related pathway can combine to modify the overall phenotype.

1.6.02 Von Hippel Lindau disease

The most common cause of inherited renal cancer is von Hippel–Lindau (VHL) disease (MIM 193300), a dominantly inherited familial cancer syndrome caused by mutations in the *VHL* tumour suppressor gene. It was named after Eugene von Hippel, who described angiomas in the eye in 1904, and Arvid Lindau, who described angiomas of the cerebellum and spine in 1927. The incidence of VHL disease is approximately 1 per 36000 live births, with similar prevalence in both genders and across all ethnic backgrounds [100, 101]. Most patients with VHL disease have a positive family history, but de novo *VHL* mutations and mosaisicm are not uncommon.

The most frequent manifestations of VHL disease are retinal and central nervous system haemangioblastomas, renal cell carcinomas, and phaeochromocytomas. These are commonly multiple and develop at a younger age than similar sporadic tumours in the general population. Patients may develop other tumours, such as endolymphatic sac tumours, and in addition to tumours, patients develop multiple cysts of the kidney and other organs including the pancreas [101-107].

Diagnosis of von Hippel-Lindau disease is based on clinical criteria or genetic testing [108, 109]. Unlike ADPKD, mutation analysis is recommended to make a definitive diagnosis and has become near universal [110]. The *VHL* gene has 3 short exons, and gives rise to a 4.5 kb mRNA with a long 3' untranslated region. Direct sequencing is the gold standard for detecting small germline *VHL* mutations [111]. "Mutation negative'' patients should subsequently be screened with techniques capable of identifying gene deletions [112]. Mutation screening should also be considered in individuals with apparently sporadic cerebral and retinal haemangioblastoma because these are rare in the general population [113-117].

Unlike ADPKD, autosomal dominant VHL disease has a pleiotropic phenotype. Early reports noted that in different pedigrees there were differing patterns of inheritance of haemangioblastoma, phaeochromocytoma, and CCRCC. This observation implied that different mutant alleles at the *VHL* locus were associated with distinct tumour suppressor capabilities [118]. There are now more than 350 distinct mutations in the *VHL* gene that have been linked to familial VHL disease [119]. The clinical phenotype is categorised on the basis of incidence of haemangioblastoma, CCRCC and phaeochromocytoma, as shown in Table 1.2. Striking aspects of the phenotype-genotype relationship are that loss of function alleles carry a

low risk of phaeochromocytoma (Type 1 disease), while some specific missense mutations predispose to phaeochromocytoma without other manifestations (Type 2C disease).

Category	Risk of Phaeochromocytoma	Risk of Haemangioblastoma	Risk of Renal Cell Carcinoma
1	Low	High	High
2A	High	High	Low
2B	High	High	High
C	Yes	No	No

Table 1.2 Classification of VHL families on the basis of tumour risk

High risk: tumour type observed in over 50% affected individuals. Low risk: tumour type observed in less than 5% of affected individuals. Yes: tumour type observed in all affected individuals. No: tumour type not observed in affected individuals.

All the clinical manifestations of VHL disease involve somatic inactivation of the wild-type *VHL* allele, in accordance with Knudson's 'two-hit' mechanism of tumour suppression [120]. This is similar to what is hypothesised to occur in ADPKD, but the resultant proliferative abnormalities lead to tumour formation not renal failure. In the kidney, the earliest, premalignant lesions in VHL patients are almost always in the distal renal tubule, suggesting that the tumours arise in this part of the nephron [121, 122]. *VHL* mutations linked to renal cancer (Type 1 and 2B) show complete loss of the ability to regulate HIF, consistent with HIF activation being critical in tumourigenesis. Although the mechanism by which HIF activation leads to tumour formation is not fully understood, it occurs as a very early genetic event

There is evidence that mutations in genes besides *VHL* can give rise to tumours similar to those seen in VHL disease [104, 123]. This may be analogous to the many ciliopathy genes than can give rise to a cystic renal phenotype [28, 124]. The *VHL* gene product has multiple roles, but the best characterised is the proteolytic degradation of the alpha subunit of the hypoxia inducible factor, HIF [125]. Mutations in other genes that affect this pathway, such as *HIF2A*, fumarate hydratase (*FH)*, or succinate dehydrogenase subunit D (*SDHD)*, have been shown to cause paraganglionomas [126] as well as other kinds of kidney cancer.

In addition to the association with VHL disease, the *VHL* gene has more recently been associated with congenital erythrocytosis syndromes. Interestingly, erythrocytosis is associated with C-terminal *VHL* mutations and it is inherited in an autosomal recessive fashion.

Moreover, mutations in *HIF2A* have also been associated with autosomal dominant familial erythrocytosis [127, 128]. These cases provide a clear link between *VHL* and systems-level response to hypoxia, showing that it plays an important role in determining a number of physiological characteristics in humans, including red blood cell production, respiratory rate and pulmonary vascular tone [129, 130].

1.7 Other mechanisms for genetic disease

The question of penetrance is important in any discussion of genetic disease. Most mutations are associated with significant phenotypic variability, both between individuals and between families. This has been shown in highly penetrant conditions, such as ADPKD, as well as in recessive conditions such as nephrotic syndrome due to homozygous mutations in phospholipase C epsilon 1 (*PLCE1*) [131]. The nature and influence of most of the modifier loci that modulate the clinical phenotypes of Mendelian traits remain unknown [132]. Conversely we know that heterozygosity for a Mendelian disorder can act as a risk factor for complex disease [133].

Digenic or oligogenic inheritance has been suggested for ciliopathies, such as Bardet-Biedel syndrome, where the clinical phenotype can be related to the overall mutational load [86]. It has also been suggested for patients with nephrotic syndrome carrying heterozygous alleles in nephrin (*NPHS1)* and podocin (*NPHS2)*, as well as for these alleles conferring an increased risk of proteinuria in the wider population [134, 135]. The modification of a single gene by one or more genes is a phenomenon known as epistasis. This interaction may explain some of the missing heritability in association studies, since the effects of interaction are often multiplicative, whereas the method for calculating heritability in GWAS is additive [58].

In addition to simple nucleotide polymorphisms, larger copy number variants represent a significant source of variation within the human genome. Re-sequencing has shown that fragments of DNA are deleted, duplicated, and inverted and that these genetic rearrangements may alter the number of copies for a particular gene [136]. We have seen how the existence of pseudo-genes is a source of mutations in *PKD1* through the process of gene conversion. Copy number variation may also play a role in the development of nephronopthisis [136]. More recently, an internal duplication in the complement factor H-

related 5 (*CFHR5)* gene was found to cause complement-mediated glomerulonephritis in the Greek Cypriot population [137]. Lastly, it has been shown that structural rearrangements of the genome increase the risk of developing complex diseases [138-140]. These structural alterations may or may not be tagged by SNPs in linkage disequilibrium, and can be hard to detect by next generation sequencing with short read lengths [59].

Alternatively, there is the possibility that some of the missing heritability unexplained by GWAS studies can be explained by mechanisms other than Mendelian chromosomal inheritance. For example, some traits can be inherited through the maternal line due to sequence changes within the mitochondrial genome; so-called extra nuclear inheritance. Mitochondrial mutations have been implicated in a range of complex diseases, including diabetes, neurocognitive disorders and even aging [141]. A number of complex mitochondrial disorders, including Leigh, MELAS and Kearns-Sayres syndromes, incorporate a renal phenotype [142]. Increased prevalence of SNPS and deletions, as well as a reduction in mitochondrial DNA copy number has been reported in patients with ESRD compared to healthy controls [24]. While the mitochondrial genome is small and easily sequenced, the issues of variable penetrance, heteroplasmy, nuclear-mitochondrial interaction, and high mutation rate of mitochondrial DNA make this area challenging to study [143].

According to latest estimates there are barely 20,000 protein-coding genes in the human genome, with even fewer disease phenotypes reported (OMIM). While protein coding sequences comprise less than 1.5% of the human DNA, current analysis suggests that 80% of the entire human genome is either transcribed, binds to regulatory proteins, or is associated with some other biochemical activity [2]. Most gene discovery using next generation sequencing has been focused on the exome, and yet the majority of SNPs identified by GWAS are intronic. While it remains technically challenging to demonstrate, it is probable that some non-coding variants will have sufficient penetrance to cause Mendelian disease.

In addition to non-coding variants, there is the possibility that variation in epigenetic regulation of gene expression may contribute to heritable renal disease. Many epigenetic studies have focused on male-female differences and in several complex diseases, including renal disease, females are relatively protected against disease progression compared to males [144]. Epigenetic regulation of X chromosomal genes, such as the angiotensin-converting-enzyme 2 gene (*ACE2*) which protects against renal damage, could be hypothesised to

influence disease outcome [145]. Similarly, changes in gene specific methylation have been reported in renal disease compared to controls [146, 147].

Lastly, some portion of heritability will be due to gene-environment interactions. These are almost impossible to study in the outbred human population, especially in the huge multinational consortia employed in large GWAS studies. One approach is to use population isolates for their genetic and environmental homogeneity. Sibling-pair studies can also reduce the effect of these factors, and are increasingly attractive as the cost of next generation sequencing continues to fall. Alternatively, inbred rodent strains provide a simplified genetic platform from which to study non-Mendelian and environmental inheritance patterns.

1.8 Conclusions

In conclusion, genetic factors play a significant role in the aetiology of kidney disease. These genetic factors exist on a continuum from rare variants that cause Mendelian disease to common variants that confer only small effect size (Figure 1.1). The major finding from renal GWAS studies is that most of the heritability of kidney disease is not accounted for by common variants (of any effect size). Consequently, this heritability must be due to the existence of many lower frequency variants. This thesis sets out to study these rare variants.

One approach to identifying causative mutations is to study population isolates that are enriched for the phenotype in question. In the third chapter I have studied the Turkish-Cypriot population of Northern Cyprus, which has a high incidence of unexplained renal disease. Familial renal disease is common in this population, and the clinical presentation of renal disease is often different to that seen in Western Europe. This suggests that novel genetic factors or disease pathways may be involved.

I have studied a number of families where multiple individuals have kidney disease and had progressed to ESRD. Technological advances have greatly simplified the identification of novel genetic variants in such cases. The question therefore becomes one of identifying the pathogenic mutation amongst a long list of private variants. This process may be achieved through the use of pedigree information, as well as bioinformatic filtering and an understanding of the biological pathways affected, as demonstrated in chapter four.

In the fifth chapter I have studied a family in which I show that kidney disease is associated with a novel mitochondrial variant. Mitochondrial disease is a source of non-chromosomal inheritance, and has recently been implicated in a range of complex diseases. The regulation of the transcription and translation of mitochondrial DNA requires a complex interaction with the nuclear genome. Moreover the mutational load and penetrance often show much greater variability than is seen in nuclear encoded genetic disorders.

Lastly, I have taken a single distinctive phenotype and related it to perturbations in the well-studied oxygen sensing pathway mediated by VHL and HIF. *VHL* mutations exhibit a high degree of allelic heterogeneity and there is a strong association between different mutations and quite different clinical phenotypes. This may reflect the ability of VHL to interact with and regulate a wide range of biological partners. It is therefore possible to use distinctive phenotypes to uncover novel cellular pathways.

Throughout this thesis I have used a variety of patient-derived cell lines to study the effect of pathogenic variants *in vitro*. Beyond the desire to explain missing heritability, these studies are the first step toward an improved understanding of biological mechanisms and the discovery of novel cellular pathways. The knowledge gained through human genetic studies will have a major impact on medical sciences. These pathways have the potential to provide new therapeutic and preventative strategies for a common and increasingly expensive worldwide health problem. Ultimately, rare variants with high enough penetrance are likely to form the basis of personalised genomic medicine in the 21st Century.

Chapter 2 Materials and Methods

2.1 Ethical approval

Collection of samples in Northern Cyprus was approved by the ethics committee of Lefkosa Burhan Nalbantoğlu State Hospital. All participants provided informed consent in writing, in accordance with the Declaration of Helsinki (Appendix A).

Collection of samples in the UK was approved by the ethics committee of the National Research Ethics Service (NRES) Committee London – West London. All participants provided informed consent in writing, in accordance with the Declaration of Helsinki (Appendix A).

Collection of fibroblasts from human volunteers was approved by Newcastle University for the MRC Neuromuscular Centre BioBank.

2.2 Recruitment of subjects

All individuals presenting with end-stage renal disease in Northern Cyprus were invited to take part in our study 'Investigation of familial renal disease in Turkish Cypriot kindreds' (chapter 3). Informed consent was obtained in writing by Dr Deren Oygar or Professor Guy Neild, and blood samples collected by those two with the assistance of Meral Yükseliş. Control DNA from the University College London Hospital haemoglobinopathy genetics service was provided by Dr Mary Petrou. 21 DNA samples were sent to me from Cyprus along with pedigree information for family 1508 (chapter 4).

Individuals were recruited in the UK in accordance with the approved protocol of the Genetic Investigation of Kidney Disease (GIKD) study. Briefly, those individuals who had already volunteered were asked to inform other family members about the study, and to ask them to contact our research team if they were interested in taking part. When these new subjects made contact, the study was explained in more detail and a time arranged to meet with them. At this subsequent meeting I would then take a full medical history and obtain informed consent to take a sample of blood or saliva for genetic analysis. I would also obtain their consent to get any other relevant medical details from their GP or hospital consultant (Appendix A).

2.3 Demographic comparisons

2.3.01 Study population

The study discussed in chapter 3 was conducted at Nicosia State Hospital in Northern Cyprus. Nicosia State Hospital is a multi-specialty tertiary care hospital, and provides renal services to the whole of Turkish administered Northern Cyprus (TRNC). All citizens of TRNC are entitled to free RRT, regardless of ethnicity. All patients with symptomatic chronic kidney disease (CKD) from within this population present either through the village practitioners, private hospitals or directly to Nicosia State Hospital. Details of every individual admitted with acute or chronic renal failure in the last decade were collected with a unique identifier number, thus preventing any duplication of records or redundancy of referrals. ESRD was defined as chronic renal disease with an eGFR of <10 ml/min/1.73 m2; but patients with diabetic nephropathy often started RRT with eGFR<15 ml/min/1.73 m2.

The most recent census of the TRNC population in 2006 showed that of 178,031 de jure citizens, 120,007 were born to parents who were both themselves born in Cyprus, providing an effective measure of the size of the ethnic Turkish-Cypriot population that has been applied throughout chapter 3 [148].

2.3.02 Variables

Basic demographic data (age at ESRD, sex, ethnic group, and probable diagnosis) were recorded for all patients receiving RRT at day 1, and again after 90 days on the renal replacement programme. Diabetic nephropathy was defined as ESRD in the presence of diabetes without evidence of an alternative diagnosis. Family history of ESRD was defined as a first or second degree relative with ESRD.

2.3.03 Data Sources for International Comparisons

European data was taken from the European Renal Association – European Dialysis and Transplant Association (ERA-EDTA) Registry report for 2008 [15]. Additional data for Greece, Turkey, and England were taken from the Hellenic Renal Registry (Dr GA Ioannidis personal communication), the National Haemodialysis, Transplantation and Nephrology Registry report of Turkey 2008 [149], and the 2008 Renal Registry Report [77]. The denominator population for international comparisons was calculated from the relevant national statistics [150, 151].

2.4 DNA Procedures

2.4.01 DNA extraction

DNA was extracted from peripheral blood, using the QIAamp DNA Blood Mini and Midi Kits (Qiagen) spin protocol, and from saliva, using the Oragene saliva kit (DNA Genotek, Ottawa, Canada). DNA was extracted from cultured cells using the QIAamp DNA Blood Mini Kit (Qiagen) spin protocol.

DNA was extracted from fresh urine after the urine was collected into 50 ml falcon tubes and centrifuged at 3000 g for 10 minutes at 4°C. The cell pellet was washed x1 in phosphate buffered saline, centrifuged again, and then re-suspended in 425 ul DNA extraction buffer (145.5 ml ddH2O, 1.5 ml Tris pH 8.0, 3 ml 1M NaCl, 1 g SDS) to which 10 ul proteinase K (10 mg/ml), 10 ul DTT (1M), and 1 ul RNAse A (25 mg/ml) had been added. Digestion was performed at 55°C for 3 hours with regular agitation.

Phenol-chloroform extraction was then performed on the digested cell pellet by the addition of an equal volume of phenol/chloroform/isoamyl alchohol (35:24:1 pH 8.0, Sigma), mixing for 5 minutes, followed by centrifugation at 130,000 rpm. After 3 such extractions, the aqueous phase was again removed and the DNA precipitated by addition of $1/10^{th}$ volume of 3M sodium acetate pH 5.2 (Fisher chemicals), 1 ul glycogen (20 ug/ul, Invitrogen), and 2 volumes of 100% ethanol. The DNA was left to precipitate at -20°C overnight, then pelleted by centrifugation at 130,000 rpm for 15 minutes at 4°C, followed by washing in 70% ethanol. After drying in air for 10 minutes, the DNA pellet was re-suspended in 100 ul Tris-EDTA buffer (Sigma).

2.4.02 Whole genome sequencing

DNA was extracted from peripheral blood, using the QIAamp DNA Blood Midi Kit (Qiagen) spin protocol. A total of 5 ug was adjusted to 75 ng/ul and integrity confirmed by running on a 0.8% agarose gel (Sigma). Samples were loaded on to a custom 96-well plate and shipped on dry ice to Complete Genomics (Complete Genomics Inc., CA). Sequence information was returned to us in the form of BAM files on a 2 terabyte hard drive.

2.4.03 Amplification Refractory Mutation System (ARMS) PCR

Gene	Direction	Temp	Sequence
COL4A3 G871C	Inner forward	68-58	CAGGTCATCAAGGTGAAATGGGACCACCGT
	Inner reverse		TCCCGGATTTCCTGGATATCCTCTTTGGCC
	Outer forward		TCGCCCCACTGCAATCCAGCCTAGAAGA
	Outer reverse		CCTGCACCATGGCCCCTGGTTTCTTACA
COL4A3 G1334E	Inner forward	68-58	GATTTCTAGGATCCATTGGACCTCCCGA
	Inner reverse		AGGTGGTCCTTTTGGCCCAATTGTTC
	Outer forward		CCCAGTGACATGCCAAAACCTATGTGAA
	Outer reverse		CCTGGGCTTCCTGGAAGGGAGATAATCT
COL4A4 3854delG	Inner forward	68-58	GCTAAAAGGGGAGAGAGGAGACCCTGTGA
	Inner reverse		AGGACCTGGAGGAGAGATTCCTGGGCTACA
	Outer forward		ATCACCACTGCAACCAGTTGTTGGTGTC
	Outer reverse		GGACCAGGTGGTCCTGAACTCCCTAAGA
GRK5	Inner forward	68-58	CTCTGCGGCTTAGAAGACCTCCAACA
	Inner reverse		TCCACTCACCGGTAGACGGTGTTCTAAC
	Outer forward		GATCCTGAGGAGGGAAGTGGAAGTGAGT
	Outer reverse		GCTTCCTGGAGTGGAAAGTCTGTCTGAC
DMBT1	Inner forward	72-63	CTGGCTCACCCACAACTGTGGCCATCAC
	Inner reverse		CACCTGAGCAGATGACACCAGCGTCTTCACT
	Outer forward		ACTTGGGTGGAGTAGGACATCACTGCTTCT
	Outer reverse		TCCTAAGTGACCACAGGAACACTAGGCAGT
INPP5F	Inner forward	65-54	AGGCTAGTCAGGAAAGAAATCAAATCAC
	Inner reverse		GATTGGGTTTCATTTGAAACTTGATAGC
	Outer forward		ATAGTAGCGTTCATGCTCCTTCAGAGAT
	Outer reverse		CTCTGAAGGAGAAGGAGTTACATGAAGC
METTL10	Inner forward	68-58	ATGAGCTCGGGCGCTGACGTCG
	Inner reverse		GCCACCGCAGCGCCACCTCT
	Outer forward		GCTCTGCCTGTGCCCCGTCTCTATGG
	Outer reverse		CCTGCCCAGCGTCACACTCACTGCT
DHX32	Inner forward	68-58	GTTGACACAGGGGATAGGTGATCAAC
	Inner reverse		TGAAAGTAAGGACATTCTACAGCAAGGAA
	Outer forward		GTCATGTATCTCCCATATCCAGCAGTTC
	Outer reverse		ACATAATGTTTGGTGTCTCTCACCACCT
PCDH15	Inner forward	65-54	ATGCCTGCCGAGAAAATTAAGTCGTC
	Inner reverse		ACACTGTCGTTGTTGATAGCTGTGTCAAAC
	Outer forward		CTCCTCCCTCTCCTCCTTCTATCCCTCT
	Outer reverse		CCCTTGTTTTGTTCAGATGTGATTTCCA

Table 2.1 Primers used for tetra-primer ARMS PCR

Primers for tetra-primer ARMS were generated using the genomic DNA sequence available (GRch37) from Ensemble (http://www.ensembl.org/Homo_sapiens/Info/Index) and the software available from the University of Southampton (http://primer1.soton.ac.uk/primer1.html) [152]. PCR analysis was performed using a G-Storm

PCR machine (Gene Technologies Ltd., UK), with a touchdown program over the temperature range and the primers indicated in Table 2.1. Products were run out on a 3% agarose gel and imaged using a bench-top UV transilluminator (UVP Ltd., UK).

2.4.04 Sanger Sequencing

The genotype of all cell cultures and specified individuals was confirmed by Sanger sequencing, following PCR amplification using the primers indicated in Table 2.2. PCR products were sequenced at the UCL Woolfson Institute for Biomedical Research using their Applied Biosystems kit, following standard manufacturer's protocols. Sequence traces were viewed on Geospiza's FinchTV and manually compared to the publically available sequence on Ensembl.

Gene	**Direction**	**Sequence**
COL4A3 exon 32	Forward	GTTAGTAGGGGAAAGCATTTGTGG
	Reverse	CTATGTACAGTTGACAGAGCCACCT
Mitochondrial D4	Forward	GCCACAGCACTTAAACACATC
	Reverse	TGCTGCGTGCTTGATGCTTG
VHL exon 2	Forward	TCCACCATGGTAGCAGAAAGGGCA
	Reverse	AGCCCAAAGTGCTTTTGAGACACC
VHL exon 3	Forward	CCTTGTACTGAGACCCTAG
	Reverse	GCTGAGATGAAACAGTGTA

Table 2.2 Primers used for Sanger sequencing

2.4.05 HSP1 sequencing

DNA samples were a kind gift of Professor Neil Sheerin. Amplification of the mitochondrial D4 region (including the HSP1 sequence) was performed using primers indicated in Table 2.2. The PCR reaction (25 µl) was set up as follows: 0.65 µl of each primer pair (20 µmol/l), 3 µl DNA (200 ng/µl), 2.5 µl 10× PCR buffer, 2.5 µl DNTPs, 2.0 µl $MgCl_2$, and 0.2 µl IMMOLASE^{TM} DNA Polymerase (5 units/µl; Bioline, UK). The reaction conditions were one cycle at 94°C for 10 min, followed by 30 cycles at 95°C for 45 sec, 60°C for 45 sec, and 72°C for 1 min, and finally 72°C for 10 min. The amplified fragments were purified using ExoSAP-IT (Affymetrix, CA) and then sequenced using a Big Dye Terminator cycle sequencing kit on an ABI Prism 377 DNA sequencer (Appied Biosystems). The sequence data were analyzed using the software

packages Factura and Sequence Navigator (Applied Biosystems), and were compared with the Cambridge sequence and MITOMAP (MITOMAP 1999).

2.4.06 Quantification of mitochondrial DNA

Relative mitochondrial DNA content was measured by expression of mitochondrial *ND1* normalised to nuclear *B2M* levels. Real-time RT–PCR analysis was performed with SYBR green (ABgene, UK) using an Opticon2 Real-time PCR machine (MJ Research, Waltham, MA) and the DNA-specific primers indicated in Table 2.3.

Gene	**Direction**	**Sequence**
B2M (genomic)	Forward	GAATGAGCGCCCGGTGTCCC
	Reverse	TCCAAGCCAGCGACGCAGTG
ND1	Forward	GCCCCAACGTTGTAGGCCCC
	Reverse	AGCTAAGGTCGGGGCGGTGA

Table 2.3 Primers used for the calculation of mitochondrial DNA copy number

2.5 Informatic analysis

2.5.01 SNP Genotyping

1 ug of genomic DNA at 75ng/ul was used for SNP genotyping. Individuals were genotyped using 300,000 single nucleotide polymorphism (SNP) markers with an average separation of 6.1 kb on the Illumina CytoSNP 12 chip (Illumina, CA, USA). Genotyping was performed at the automated, high-throughput system available at UCL Genomics.

2.5.02 Linkage analysis

Linkage analysis was performed on pedigrees using the easyLINKAGE Plus v5.05 graphical user interface for two-/multi-point linkage analysis [153], with analysis by GeneHunter [154] and SimWalk [155]. Pedigrees were confirmed using HaploPainter V.1.043 [156] and PedCheck

[157]. For calculation of parametric LOD scores a 'rare dominant' model was utilised, with disease allele frequency set to 10-4 and disease penetrance vector set to 1.0.

2.5.03 Calculation of the minimal shared haplotype

The minimal haplotype shared by 4 discrete pedigrees with the G871C mutation was determined using SNP genotype data. Conflicting homozygosity was assessed in this genotype data, using a program written by Dr Daniel Gale and Dr Adam Levine at UCL. I then determined the relative position of the boundary SNPs in centimorgans.

2.5.04 SNP variant filtering

The Complete Genomics data was filtered using CGAtools, annotated and filtered again using Excel. Variations in the 'All variations' from the reference human genome at the linked loci was extracted from the master variants file using CGAtools (http://cgatools.sourceforge.net/). These variants were then annotated using Annovar (http://www.openbioinformatics.org/annovar/) and filtered to exclude variants that were intergenic; intronic (but not including splice site); synonymous or common (>0.5% allele frequency in 1000 genomes database). The remaining shared variants that are predicted to change amino acid sequences were considered candidate mutations.

2.5.05 Analysis of pathogenicity

Variants identified by whole genome sequencing (WGS) were analysed for pathogenicity using PyloP, Mutation taster, SIFT, and PolyPhen2. Evolutionary conservation was assessed using the UCSC table browser tool to generate PhyloP scores across 46 vertebrate species (http://genome.ucsc.edu/cgi-bin/hgGateway). The predicted effect of an amino-acid exchange was assessed using MutationTaster (http://www.mutationtaster.org/) [158], SIFT (http://sift.bii.a-star.edu.sg/) [159], and PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/) [160].

2.6 Tissue Culture procedures

2.6.01 Creation of primary cell lines

2.6.01.01 Dermal fibroblasts

All fibroblast lines were created at the Rayne Institute at UCL. 4mm skin punch biopsies were performed under local anaesthesia using aseptic technique in the patient's own home. Skin wounds were secured with steri-strips and dressed. The biopsy tissue was stored in complete DMEM supplemented with penicillin/streptomycin at room temperature. Biopsies were collected from affected patients and normal controls on the same day and processed in parallel.

Skin biopsies were divided into two equal portions. One portion was digested with 2 ml pre-warmed 2.5% (10x) Trypsin (Sigma) for 15 minutes at 37°C with occasional agitation in a 15 ml Falcon tube. This sample was then centrifuged at 400 g for 5 minutes at room temperature, and the supernatant removed. The tissue was further digested with 2 ml pre-warmed collagenase at 1 mg/ml (Sigma, C1889) at 37°C with occasional agitation. After 90 minutes the collagenase was neutralised with 10 ml warmed complete Ham's F10 media, and the samples centrifuged. The samples were washed once with 10 ml complete media, before being plated out in T25 flasks in 5 ml Ham's F10 media (Sigma), supplemented with 10% foetal bovine serum (FBS), 5mM of L-glutamine (Gibco), penicillin/streptomycin antibiotics (Gibco), 1 mM pyruvate (Gibco), 50 ug/ml uridine (Sigma), 2.5 ug/ml amphotericin B (Sigma), and 50 ug/ml gentamicin (Sigma) (Complete media). The other portion was cut into small pieces with a scalpel and fixed to a T25 flask by air drying for 15-30 minutes. 4 ml complete Ham's F10 media was then added to the flask, taking care not to re-suspend the pieces of tissue.

Following satisfactory growth, cells were maintained in DMEM cell culture medium (Invitrogen), supplemented with 10% foetal bovine serum (FBS), 5mM of L-glutamine (Gibco), penicillin/streptomycin antibiotics (Gibco), 1 mM pyruvate (Gibco), and 50 ug/ml uridine (Sigma). Fibroblasts were cultured in standard culture flasks (Nunc) at sub confluent density, split 1:3 – 1:5 twice a week, and kept in a humidified incubator at 37°C gassed with ambient air plus 5% CO2. All lines were confirmed free of mycoplasma infection.

Cells were treated with ethidium bromide (Sigma-Aldrich, St Louis, MO) at 50 ng/ml for a minimum of 1 week to produce 'rho zero' control cells.

2.6.01.02 Urothelial cells

Urothelial cell lines were cultured according to established protocols [161]. Briefly, urine samples were collected in sterile 500 ml glass bottles in the patient's own homes and transferred to the laboratory at room temperature. They were then divided into 50 ml aliquots inside a tissue culture hood and centrifuged at 400 g for 10 minutes at room temperature. The supernatant was discarded and the cell pellets pooled and washed in 37°C phosphate buffered saline (Gibco) containing 2.5 ug/ml amphotericin B (Sigma), 50 ug/ml gentamicin (Sigma) and penicillin/streptomycin. This was centrifuged at 400 g for 10 minutes at room temperature, and washed a second time, if the pellet was especially large. The supernatant was again discarded, and the remaining 0.2 ml of sample was re-suspended in primary medium.

Urothelial cell primary medium contained a 1:1 mix DMEM (Gibco) and Ham´s F12 (Sigma), with 10 % fetal bovine serum, REGM™ SingleQuots™ Kit (Lonza), 2.5 ug/ml amphotericin B (Sigma), 50 ug/ml gentamicin (Sigma), and penicillin/streptomycin. The cell suspension was divided between the wells of a 12-well plate that had been coated with L-gelatine (Gibco). Over the next few days a few add hundred µl of primary medium was added to each well. Subsequently the primary medium was changed to Renal Epithelial Basal Medium (REBM™, Lonza, Cologne) to which SingleQuots™ supplements and growth factors (hydrocortisone, hEGF, FBS, epinephrine, insulin, triiodothyronine, transferrin and gentamicin/amphotericin-B; Lonza, Cologne) had been added as per the manufacturer's instructions.

Visible colonies appeared routinely after 3-6 days. After visual inspection to confirm epithelial morphology, cells were mobilised when confluent with 0.25% trypsin containing 1 mM EDTA. Urothelial cells were cultured in standard culture 6-well plates (Nunc) coated with L-gelatine (Gibco) and harvested after 1 or 2 passages.

2.6.01.03 Lymphoblastoid cells

The first lymphoblastoid cell lines (LCL) were created at the Institute of Child Health at UCL by Dr Jasper de Boer. Blood was collected from the index patient (Mt1) and normal controls. Control lymphocytes from four healthy donors were pooled to give four control lines (C1, C2, C5, C6) thus containing the genetic information from 12 individuals. The second LCL lines were

created at the Rayne Institute at UCL by me, using blood collected from the index patient (Mt2) and two further normal controls (C3, and C4).

Mononuclear cells were isolated from peripheral blood by Ficoll gradient centrifugation and re-suspended in RPMI-1640 culture medium with 20% fetal calf serum (FCS), 2mM L-Glutamine and 1% Penicillin-Streptomycin. Epstein Barr virus (EBV) supernatant was added, followed 24 hours later by cyclosporine A (10µg/ ml), with acyclovir (100µM) administered 14 days later.

Following satisfactory growth, cells were maintained in RPMI 1640 cell culture medium (Invitrogen), supplemented with 10% fetal bovine serum (FBS), 5mM of L-glutamine (Gibco), and penicillin/streptomycin antibiotics (Gibco). Lymphocytes were cultured in standard culture flasks (Nunc) at a density of $1x10^6$/ml, and kept in a humidified incubator at 37°C gassed with ambient air plus 5% CO2.

2.6.02 Cell culture techniques

2.6.02.01 Seahorse analysis

Cells were plated at in the proprietary XF24 cell culture micro-plates 24 hours before analysis (Seahorse bioscience, MA). After counting, the cells were left to attach in 100 ul medium for 5 hours, before this volume was made up to 250 ul. Analysis plates were calibrated with 1 ml XF calibrant fluid at 37°C in a CO2-free incubator for a minimum of 4 hours prior to use. Immediately prior to analysis, the culture media was aspirated, the cells were washed with 1 ml seahorse assay media (Bicarbonate-free DMEM, Sigma, made up as for standard culture with antibiotics, glutamine, pyruvate and uridine), and the wash was replaced with 600 ul seahorse assay media.

The measurement of mitochondrial function was performed with the standard mix-wait-measure cycle of 3 minutes-2 minutes-3 minutes. 4 untreated baseline cycles were followed by injection of oligomycin, 3 cycles then injection of FCCP, 3 cycles then injection of actinomycin, then a final 3 cycles. Compounds were added at the concentrations indicated below.

2.6.02.02 Creation of RCC-10 cell lines containing VHL mutant constructs

The VHL mutant constructs were created by Dr Daniel Gale at UCL. Point mutants were generated in wild-type VHL cDNA in the pCMVR vector using Quikchange (Stratagene). Each mutant was confirmed by sequencing.

VHL-null RCC-10 cells [162] cells were transfected by Dr Daniel Gale with 2µg of plasmid DNA using Lipofectamine 2000 (Invitrogen). Stably transfected cell pools were obtained by geneticin (G418, Invitrogen) selection.

2.6.02.03 Experimental reagents

Hypoxic experiments were performed using a humidified hypoxic workstation (InVivo2 1000, Ruskinn technology, Cardiff, UK) with 1% ambient 02 and 5% CO2 for 16 hours. Cells were treated with the reagents listed below for 16 hours to stabilise HIF-1a Table 2.4. Cells were treated with proteasomal and protease inhibitors listed below for 3 hours.

Reagent	**Company**	**Working concentration**
Oligomycin	Sigma-Aldrich	5 ug/ml
FCCP	Sigma-Aldrich	7.5 ug/ml
Antimycin A	Sigma-Aldrich	4 ug/ml
Cycloheximide	Sigma-Aldrich	50 ug/ml
Desferrioxamine (DFO)	Sigma-Aldrich	25 uM for 16 hours
Dimethyloxalylglycine (DMOG)	Frontier Scientific, Logan, UT	1 mM for 16 hours
MG-132	Merck Chemicals Ltd, UK	10 uM for 3 hours
ALLN	Roche Molecular Biochemicals	50 uM for 3 hours
MG-115	Calbiochem, UK	30 uM for 3 hours
PD150606	Calbiochem, UK	20 uM for 3 hours
Caspase 3 Inhibitor II	Calbiochem, UK	50 uM for 3 hours
Cathepsin Inhibitor I	Calbiochem, UK	50 uM for 3 hours

Table 2.4 Compounds used in cell culture experiments

2.6.02.04 Lentiviral-mediated knockdown

Lentiviral transduction particles were obtained from the MISSION shRNA Library (Sigma Aldrich). These particles are produced from a library of sequence-verified lentiviral plasmid shRNA cloned into the pLKO.1-puro vector, and express shRNA on the U6 viral promoter. LCL cells were infected at an MOI of 1 in the presence of polybrene (hexadimethrine bromide) at 6 ug/ml (Sigma-Aldrich, St Louis, MO), and then spun at 1800 RPM for 45 minutes at 25 degrees immediately following addition of virus. Infected cell pools were selected for puromycin resistance (Sigma-Aldrich), and expanded over 3-4 weeks until suitable numbers were obtained.

Lentiviral particles derived from five VHL-specific shRNA were assessed by RT-PCR and western blot and the one with greatest knockdown was selected. This shRNA targeted the end of VHL exon 2, and contained the sequence:

5'-CCGGTATCACACTGCCAGTGTATACCTCGAGGTATACACTGGCAGTGTGATATTTT TG-3'.

The control lentiviral particles were derived from the same pLKO.1-puro vector and contained an shRNA sequence targeting Luciferase, Photinus pyralis (GenBank Accession No. M15077). The Luciferase shRNA Control Particles are useful as a positive knockdown control in experiments using cell lines expressing firefly luciferase but were used in this case as a nonsense control since LCL cells do not contain this enzyme.

2.7 Western Blotting

Cellular extracts were harvested using 8M urea buffer with 1/100th volume PMSF (Fluka) [125]. Protein levels were quantified using BCA Protein Assay (Pierce) measuring absorbance at 560nm on a Tecan GENios fluorescence reader. Equal amounts of protein were mixed with 1/8th volume loading buffer (12.5 ml 0.5M Tris pH 6.8 (Sigma), 10 ml glycerol (Sigma), 2 g sodium dodecyl sulphate (Sigma), 1.55 g dithiothreitol (Fluka), 0.25 mg bromophenol blue (Sigma), with deionised H20 to 50 ml). Proteins were resolved on a polyacrylamide gel and transferred to a polyvinylidene fluoride (PVDF) membrane (Immobilon-P, Millipore). Molecular weight standards (BioRad) were used to calibrate the molecular weights of proteins. After transferring, the membrane was incubated for one hour with 5% skimmed milk in Tris-buffered saline containing 1% Tween 20 (Sigma) (TBS-T). The membrane was then incubated with the primary antibody of interest, indicated in Table 2.5. Then it was washed three times in TBS-T and incubated with the horseradish peroxidase (HRP) conjugated secondary antibody for 1 hour at room temperature. After a further 3 washes in TBS-T, immunoreactivity was detected by chemiluminescence (ECL, GE Healthcare). Films were developed with a Photon SRX-101a film processor (Konica-Minolta).

Antibody	Origin	Company	Concentration
VHL Ig32	Mouse	BD Pharmigen	1:1000
HIF-1a	Mouse	BD Pharmigen	1:1000
HIF-2a	Rabbit	Abcam	1:1000
Tubulin	Mouse	Sigma	1:5000
ß-actin	Mouse	Abcam	1:10,000
GLUT-1	Rabbit	Alpha Diagnostics	1:1000
MT-CO1	Mouse	Abcam	1:1000
VDAC1	Mouse	Abcam	1:1000
Cytochrome C	Mouse	BD Biosciences	1:1000
Anti-mouse HRP	Sheep	GE Healthcare	1:5000
Anti-rabbit HRP	Goat	Santa Cruz	1:5000

Table 2.5 Antibodies used in western blotting

2.8 Real-time RT-PCR

Total RNA was extracted from cells using RNABee (Biogenesis). For the analysis of mitochondrial transcripts, RNA was then subjected to DNase digestion (Qiagen), followed by RNA cleanup using the RNeasy spin column (Qiagen). 0.5 - 1 ug of total RNA was retro-transcribed using a first strand cDNA synthesis kit (Quanta Biosciences) according to the manufacturer's instructions.

Real-time PCR analysis was performed using SYBR green (ABgene, UK), using an Opticon2 Real-time PCR machine (MJ Research, Waltham, MA) and the primers indicated in Table 2.6. The PCR reactions were performed with the following cycling parameters: 95°C for 5 minutes; 40 cycles of 94°C for 20 seconds, 60°C for 25 seconds, 72°C for 50 seconds; followed by melting curve analysis 65°C - 94°C. All real-time PCR analysis results presented were normalised to levels of housekeeping gene expression.

Gene	**Direction**	**Sequence**
β-Actin	Forward	CCCAGAGCAAGAGAGG
	Reverse	GTCCAGACGCAGGATG
E-cadherin	Forward	GACAACAAGCCCGAATT
	Reverse	GGAAACTCTCTCGGTCCA
KRT7	Forward	CTGCCACGCCCACTTCCCAG
	Reverse	CAGGCGCGGCTTCCTGTCAA
L1CAM	Forward	AGTTGGAGGCCGCCATTGCC
	Reverse	CCAGGGCCAGCTTCACGCTC
INPP5F	Forward	TCAGAGTGCACCCTCATTGATGC
	Reverse	AGGCCACGTAGTAGGCAGAGTT
DHX32	Forward	ACCCCTTTGATGGATTGCCA
	Reverse	TGAGGAACCTGAGCGCTCTTAC
PCDH15	Forward	AACCGCAAATCAGCTTTGAGTGAAC
	Reverse	ACTGGCCCAAGCAGATTTCAAAGA
B2M (RNA)	Forward	GTGCTCGCGCTACTCTCTCT
	Reverse	TCAATGTCGGATGGATGAAA
MT-CO1	Forward	ACGTTGTAGCCCACTTCCAC
	Reverse	CATCGGGGTAGTCCGAGTAA

RNR1	Forward	CTAGCCACACCCCCACGGGA
	Reverse	CGCGGTGGCTGGCACGAAAT
ND6	Forward	AGGTAGGATTGGTGCTGTGG
	Reverse	CCAATAGGATCCTCCCGAAT
VEGF	Forward	TGCCAAGTGGTCCCAG
	Reverse	GTGAGGTTTGATCCGC
P4HA1	Forward	GCGACTGGGGGCTGAAGAGC
	Reverse	GCCTGGATGAGCCAAAGACTGGG
PFKFB4	Forward	CCAGGACTGTCCCCTCGGGG
	Reverse	CACACCCAGTGCCTCAGCCG
VHL	Forward	TGCCATCTCTCAATGTTGA
	Reverse	CTCCGGACAACCTGGA

Table 2.6 Primers used for gene expression analysis

In the absence of previously published sequences, primers were designed using the nucleotide sequence (http://www.ncbi.nlm.nih.gov/nuccore) and the NCBI programme Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). They were designed to span an exon-exon junction to confer additional specificity where possible. Primers were subsequently validated on sequential dilutions of template cDNA in order to determine efficiency of amplification, melt curve analysis, and assessment of PCR product by agarose gel electrophoresis.

2.9 Analyses of data

2.9.01 Population statistics

The incidence of RRT in the Turkish-Cypriot population was averaged over eight consecutive calendar years (2004–2011). Age-standardised incidence rates were calculated using the Eurostat EU27 population figures [163]. All comparisons with other populations were made using data for chronic RRT; that is those still on RRT at 90 days. The emphasis of this report is on age-specific comparisons as these are not affected by age-referral biases. 95% confidence intervals were calculated, assuming a Poisson distribution for incidence rates. Statistical analysis was carried out using Stata version 11.

2.9.02 Calculation of target mRNA transcript expression

Target gene expression was calculated following the determination of the Ct values for each experimental sample. A relative method of quantification was used as described by Pfaffl [164], with actin as the internal reference gene by equation 1:

Equation 1: relative target gene expression = ΔCt target/ΔCt reference

In this equation, ΔCt target represents the difference in Ct values of the target gene between the calibrator sample and the experimental sample. Similarly, ΔCt reference represents the difference in Ct values of the reference gene between the calibrator sample and the experimental sample.

2.9.03 Analysis of data

Once relative expression had been determined in all of the experimental samples, the results were formatted into the statistical software package GraphPad Prism version 4 (San Diego, CA). Data were analysed using repeated measures ANOVA.

Chapter 3 Population genetics

3.1 Introduction

3.1.01 The genetics of island populations

Renal disease is more common in certain racial and ethnic groups, which reflects a complex interaction of socioeconomic, environmental, and genetic factors [14, 15, 77, 165, 166]. These genetic factors are determined by differences in allelic frequencies between population groups. Allelic frequencies are stable in large, isolated populations, where mating is random. Genetic equilibrium is maintained only in the absence of any appreciable rate of mutation, the absence of selection against any particular genotype, and no migration. Factors that disturb these assumptions will result in changes to allelic frequency, and this is particularly seen in island populations.

Island populations can be small, and therefore susceptible to the process of genetic drift. Genetic drift describes the change in frequency of a genetic variant due to chance events. Such chance events affect the reproductive fitness of carriers of a mutation independently of the effect of the given mutation. When there are few copies of an allele, the effect of genetic drift is larger. Moreover, populations derived from a small number of individuals as a consequence of some type of bottleneck will show reduced allelic diversity. This is important because in large populations, high allelic diversity may significantly reduce the contribution of risk alleles.

Reduced allelic diversity may also arise from non-random mating. In human populations, non-random mating may occur for a number of reasons: stratification, assortative mating, and consanguinity. Stratification describes a population in which there are a number of subgroups that have remained genetically separate during modern times. Assortative mating is often positive, and describes the choice of a mate because that individual possesses some particular trait, such as native language or skin colour. Consanguinity also acts to increase the frequency of autosomal recessive disease

Migration can change the allele frequency by the process of gene flow, defined as the slow diffusion of genes across a reproductive barrier. In contrast to the reasons given for non-random mating, gene flow usually involves a large population and a gradual change in gene frequencies. In human populations, the barrier may be racial, ethnic, or cultural and not necessarily geographical requiring physical movement from one region to another.

3.1.02 Genetic disease on the island of Cyprus

The island population of Cyprus is an interesting example for the renal geneticist. Cyprus is an island country in the eastern Mediterranean that has been occupied by a series of historical powers, resulting in waves of successive migration. The island was first settled in the 10th millennium BC by people from the Levant, and has experienced long periods of Greek rule from the 2nd millennium BC, and more recently Turkish rule following the Ottoman invasion of the island in 1571. Genetic analysis of the Cypriot population suggests that, despite close historical ties, it is different to that of either Greece or Turkey [167, 168]. The distribution of thalassaemia alleles within the island also suggests that despite four centuries of cultural separation, the Greek and Turkish Cypriot communities show significant gene flow [169, 170].

Turkish Cypriots form a distinct ethno-linguistic community centered on Turkish administered Northern Cyprus (TRNC). The most recent census of this population in 2006 showed that 120,007 *de jure* citizens were born to parents who were both themselves born in Cyprus, providing an effective measure of the size of the ethnic Turkish-Cypriot population [148]. Recently, a number of inherited renal diseases have been described in the Greek Cypriot population [137, 171]. Given the evidence for significant gene flow between communities, it is probable that these conditions also exist in the Turkish-Cypriot population.

Several mutations have shown significant geographical clustering in the Greek population of Cyprus, affecting medullary cystic kidney disease 1 (*MCKD1)*, frataxin (*FXN)*, pyrin (*MEFV)*, and polycystin 2 (*PKD2)* [81, 172-174]. This may be due to the widespread practice of marrying within one's close local geographical area, which is reflected in the well-known old village Cypriot saying:

Παπούτσιν που του τόπου σου τζιας ευ κομμαδκιασμέυου.

'you had better obtain shoes from your own place, even though they may be full of holes' [173].

This form of non-random mating is distinct from the consanguineous intermarriage seen in some cultures, but has a similar, although lesser, effect on perpetuating recessive and dominant mutations.

Complement factor H-related 5 (CFHR5) nephropathy, which affects between 1 in 1000 and 1 in 8000 Greek-Cypriots, is caused by an internal duplication in the *CFHR5* gene [137]. It is characterised by microscopic haematuria, renal failure, and C3 glomerulonephritis. The extent of the shared haplotype in affected individuals suggests that the mutation arose in a common ancestor about 16 generations ago. It is highly penetrant, and may be a significant cause of renal disease affecting inhabitants of the island and their descendants worldwide.

Three specific mutations in collagen, type-IV, alpha-3 (*COL4A43)* and collagen, type-IV, alpha-4 (*COL4A4)* show a similar founder effect in the Greek-Cypriot population [171]. These mutations were identified in 13 pedigrees that co-segregated microscopic haematuria, mild proteinuria, and variable degrees of renal impairment. Renal biopsy gave a dual diagnosis of thin basement membrane nephropathy (TBMN) and focal segmental glomerulosclerosis (FSGS). Whilst TBMN is often considered a benign condition, these mutations are associated with a significant risk of progression to end-stage renal disease. Thus 19.6% of heterozygotes showed progression to ESRD over three decades of follow-up, and that proportion was increased to 46% in those aged over 51 years [171].

Given the selective pressure operating against pathogenic variants, single gene disorders are usually of recent evolutionary origin [78]. The population arising after a genetic bottleneck will exhibit reduced genetic diversity, due to the presence of a small number of founders. In isolated populations, it is probable that all individuals carrying a given mutation will share a common founder. It is possible to look at the ancestral haplotype with genetic markers to determine whether the variant was indeed shared by a common ancestor, and thereby to determine the age of the mutation.

Single gene disorders are usually rare in the general population, and therefore unlikely to be a significant contributor to the prevalence of renal disease. However, isolated populations that exhibit assortative mating may show increased frequency genetic disorders. Such populations show increased frequencies of homozygosity, and consequently of recessive disease, as seen in the Finnish population and Old Order Amish and Hutterite communities [175]. The effect of selection pressure to remove these deleterious alleles may also be reduced if the disorder is manifest after childbearing age.

3.1.03 Renal disease in Cyprus

Clinical observation suggested that the Turkish Cypriot population of Northern Cyprus exhibited a high incidence of renal disease. We therefore designed a study to quantify the incidence and prevalence of renal disease, as measured by those on renal replacement therapy (RRT), in this ethnically-defined population disease [176]. The collection of accurate epidemiological data is of great importance for healthcare policy planning [14, 15, 77] and this is especially true for places where the infrastructure delivering RRT is improving, such as the island of Cyprus. To put these data into context I examined comparable incidence data from Greece, Turkey, and the white population of England.

The incidence of RRT is a complex end-point. Renal disease has a wide range of aetiologies, and survival to ESRD and the provision of RRT is influence by a wide range of factors [9, 177]. There are considerable differences in the incidence and prevalence of RRT within Europe. Registry data demonstrate a North-South gradient, with higher incidence of RRT and lower mortality around the Mediterranean [15, 178]. Importantly, data from the Eastern Mediterranean is more difficult to obtain, and information on primary renal disease is less robust [179]. In order to understand the aetiology of ESRD and chronic kidney disease, which affects many more people, it is important to determine the primary renal diagnosis [179]. ESRD attributed to type 2 diabetes and hypertension continues to rise throughout the world, and this is increasingly true for countries such as Cyprus [15, 180-182].

An important, and probably unique, aspect of this study is that every single patient on RRT provided a DNA sample for research. Because the Greek and Turkish communities share many genetic characteristics [169, 170], I initially undertook genetic testing for two conditions that have been identified in Greek Cypriots, CFHR5 nephropathy and thin basement membrane nephropathy due to the three mutations in *COL4A3* and COL4A4 outlined above [137, 171]. The aim of this analysis was to evaluate the contribution of these conditions to RRT incidence. The discovery of these conditions within the Turkish Cypriot population would further support the existence of gene flow between these populations and it would enable more accurate determination of the age of these mutations.

3.2 Results

A total of 225 Turkish-Cypriot patients were maintained on renal replacement therapy beyond 90 days during the study period (01/01/04 – 31/12/11). More males than females started RRT (139 vs 86), but the age distribution was similar between the sexes (Table 3.1). The average crude and age-standardised incidence rates at day 1 for this period were 311.4 and 456.9 per million population, respectively. The Turkish-Cypriot case mix at 90 days included 36.0% with ESRD due to diabetic nephropathy, 29.8% with unknown diagnosis, and 3.6% with polycystic kidney disease (Table 3.2).

Characteristic	Category	Turkish-Cypriot		Turkish minority	
		Number	Percent	Number	Percent
Gender	Male	139	61.7%	43	64.2%
	Female	86	38.3%	24	35.8%
Median age of adult population at ESRD in years (lower and upper quartile)	Male	63 (54,69)		51 (39,61)	
	Female	64 (52,74)		58 (49,66)	
Modality of RRT at presentation	Haemodialysis	205	91.1%	60	89.5%
	Peritoneal dialysis	10	4.4%	5	7.5%
	Renal transplant	8	3.5%	0	0%
	Unknown	0	0%	2	3.0%
Co-morbidities	Diabetes	93	41.3%	22	32.8%
	Hypertension	145	64.4%	39	58.2%
	Family History	41	18.2%	5	7.5%

Table 3.1 Baseline characteristics of the incident Turkish-Cypriot renal replacement therapy (RRT) population at 90 days

National registry data shows that Turkey, Greece, and Tunisia have an age-adjusted incidence rate at day 90 of 349.1, 159.5, and 238.7 pmp respectively [15]. In comparison the crude and age-adjusted incidence rates of chronic RRT for Turkish-Cypriots were 234.4 and 315.0 per million population, respectively. Figure 3.1 shows that the incidence of RRT in each age group in Turkish-Cypriots is comparable to that seen in Turkey. Despite the small number of cases, the incidence of RRT in the 45-64 age group in Turkish-Cypriots (412.0 pmp) is significantly higher than that seen in the white population of England (123.2 pmp).

Diagnosis	**Turkish-Cypriots**		**Greece**		**Turkey**		**England whites**	
	pmp	**%**	**pmp**	**%**	**pmp**	**%**	**pmp**	**%**
Diabetes	84.4	36.0%	53.5	30.3%	61.6	29.9%	24.5	19.5%
Glomerulonephritis	29.2*	12.4%*	14.8	8.4%	16.4	7.9%	13.5	10.7%
Hypertension	1.0	0.4%	19.1	10.8%	53.4	25.9%	6.8	5.4%
Renal Vascular Disease	12.5	5.3%	3.0	1.7%	2.8	1.3%	7.8	6.2%
Data not available	0.0	0.0%	0.0	0.0%	3.1	1.5%	9.8	7.8%
Other identified Category	12.5	5.3%	15.7	8.9%	21.0	10.2%	18.2	14.5%
Pyelonephritis	16.7	7.1%	9.9	5.6%	7.6	3.7%	9.7	7.7%
Polycystic Kidney	8.3	3.6%	6.7	3.8%	7.8	3.8%	10.1	8.0%
Uncertain aetiology	69.8	29.8%	53.9	30.5%	32.5	15.7%	25.5	20.3%
Total incidence at day 91	234.4	100.0%	176.7	100.0%	206.2	100.0%	125.9	100.0%

*Includes presumed glomerulonephritis not biopsy proven.

Table 3.2 Provisional renal diagnosis in the Turkish-Cypriot renal replacement therapy (RRT) population

Incidence at 90 days, expressed as incidence per million population and percentage, in comparison with 2008 registry data for Greece, Turkey, and the UK.

There is a high rate of RRT for ESRD attributed to diabetic nephropathy in all eastern Mediterranean countries (84.4 pmp in Turkish-Cypriots, 53.5 pmp in Greece, and 61.6 pmp in Turkey) (Table 3.2). The code 'uncertain aetiology' is more common in the Turkish-Cypriot registry (69.8 pmp) than in other populations. However there were also lower levels of hypertensive nephropathy in Turkish-Cypriots than in Greece or Turkey (1.0 pmp vs 19.1 pmp and 53.4 pmp respectively). The incidence of polycystic kidney disease was broadly similar across all registries (range 6.7 - 10.1 pmp).

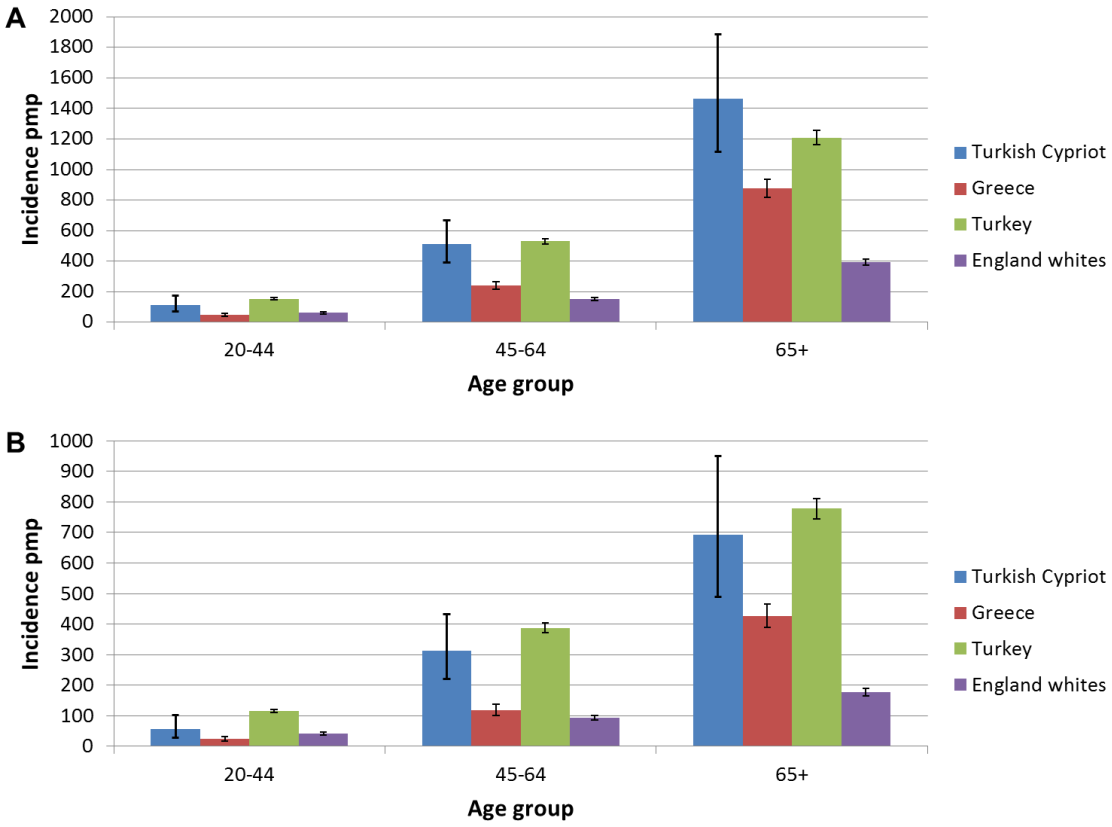


Figure 3.1 Incidence of renal replacement therapy (RRT) at 90 days by age and gender

Incidence of RRT at 90 days in Turkish Cypriots compared with 2008 registry data for Greece, Turkey, and English whites in males (A), and females (B). Error bars indicate 95% confidence intervals.

The incidence of RRT in the minority Turkish population of Northern Cyprus is broadly in line with the rate seen in Turkish-Cypriots (Table 3.1). An additional 67 Turkish patients were maintained on renal replacement therapy beyond 90 days during the study period, giving an overall incidence of 211.3 pmp in the *de jure* population of Northern Cyprus. However data on this population is less accurate due to significant migration to and from the mainland of Turkey [148].

Figure 3.2 shows the percentage of the incident population with a family history of ESRD by age and patient group. 18.2% of the incident population had a first or second degree relative with ESRD; this rose to 37% in those with ESRD due to uncertain aetiology. Interestingly, if we consider all those with a broadly glomerular or tubular phenotype, then the proportion with a family history becomes 62.5%.

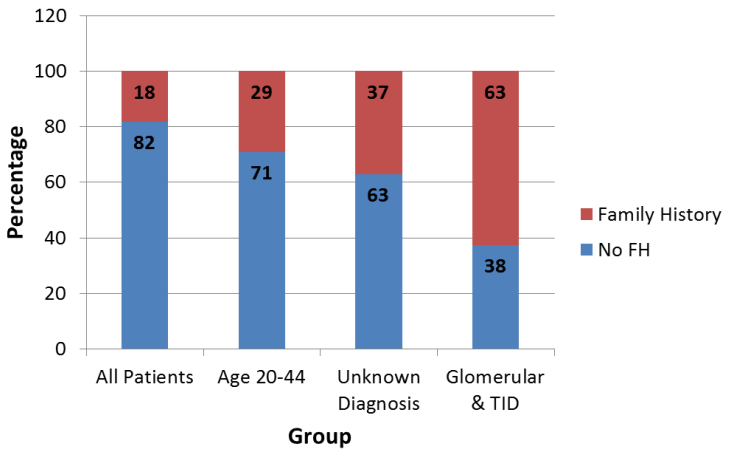


Figure 3.2 Family history of ESRD in different patient groups

Family history is especially common in those with unknown diagnosis, and those presenting with a glomerular or tubular phenotype [179].

DNA was collected from the entire Turkish Cypriot population on RRT. Because the Greek and Turkish communities share many genetic characteristics [169, 170], we initially undertook genetic testing for two conditions that have been identified in Greek Cypriots, CFHR5 nephropathy and thin basement membrane nephropathy [137, 171]. Table 3.3 shows the results of tetra primer PCR in the Turkish Cypriot population with ESRD and a control Turkish Cypriot population. Three individuals demonstrated the G871C mutation in *COL4A3* previously reported in Greek-Cypriot pedigrees [171]. I did not detect any of the previously cited mutations in *COL4A4* or the *CFHR5* duplication in the ESRD population [137, 171].

	ESRD		**Control**	
Mutation	Cases	MAF	Cases	MAF
COL4A3 (G871C)	3	0.013	0	0.000
COL4A3 (G1334E)	0	0.000	0	0.000
COL4A4 (3854delG)	0	0.000	0	0.000
CFHR5 duplication	0	0.000	0	0.000

Table 3.3 PCR analysis of mutations in COL4A3, COL4A4, and CFHR5 in the Turkish-Cypriot population

A total of 225 Turkish-Cypriot patients with ESRD and 76 control patients (from the University College London Hospitals haemoglobinopathy genetics service) were genotyped using tetra-primer ARMS PCR. 3individuals were found to carry the G871C mutation in *COL4A3*, giving a minor allele frequency (MAF) of 0.013.

I also examined a number of Cypriot families based in London that showed evidence of inherited renal disease. One family, 1509, consisted of two sisters who were currently on dialysis with diagnoses of diabetic nephropathy and analgaesic nephropathy (Figure 3.3A). One son (442) was noted to have haematuria and proteinuria whilst he was being worked up as a possible live donor for his mother. Subsequent renal biopsy showed TBMN on biopsy, as shown in Figure 3.3B. His cousin (401), living in Australia, also had haematuria and proteinuria. Tetra primer PCR and Sanger sequencing confirmed the presence of the G871C mutation in all those clinically affected with renal disease (Figure 3.3C and D).

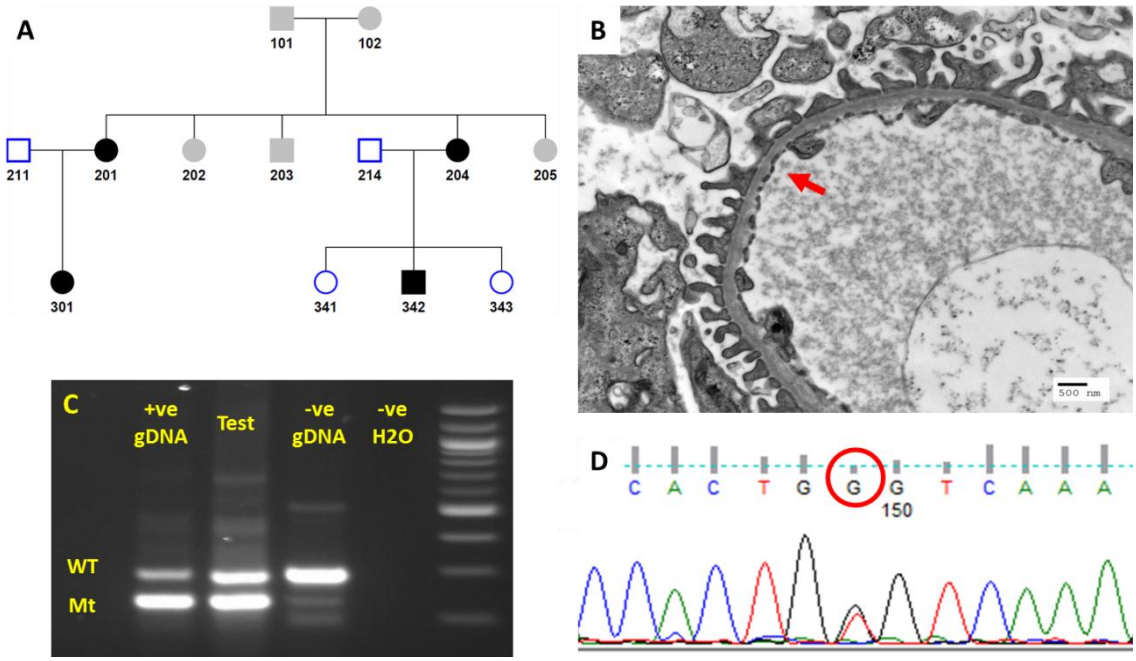


Figure 3.3 Family 1509

(A) Pedigree of family 1509 showing carriers of the G871C mutation in *COL4A3*. (B) Representative EM image from an affected family member showing thinning of the GBM (arrowed). (C) Tetra primer PCR showing amplification of mutant (Mt) and wild-type (WT) allele in individual 342 (Test) compared to one of the original Greek Cypriots with this mutation (+ve) or control DNA (-ve). (D) Sequence trace demonstrating heterozygous c.2512G>T (p.Gly871Cys) in *COL4A3* in individual 342.

Two individuals who carried the G871C mutation in the Turkish-Cypriot ESRD population were noted to come from a large pedigree that we had recruited at the same time as collecting samples from the ESRD population. This family, termed 1505 is shown in Figure 3.4. Family 1505 co-segregated microscopic haematuria, mild proteinuria, and renal impairment, however initial linkage analysis failed to show association with a single genetic locus. Subsequent testing for the G871C mutation in this family revealed 15 heterozygote carriers, and the presence of four phenocopies that had confounded the linkage analysis.

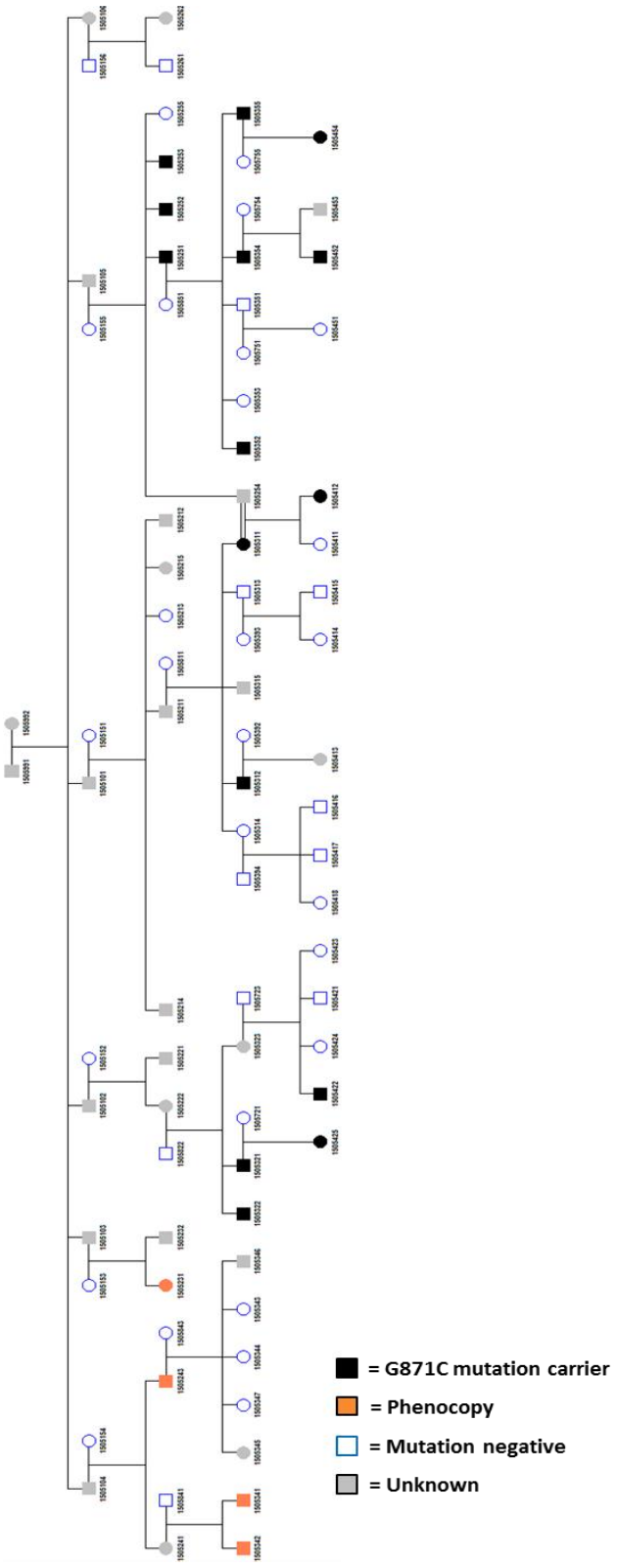


Figure 3.4 Pedigree of family 1505

The family includes 15 carriers of the G871C mutation in *COL4A3* and 4 mutation negative phenocopies with evidence of renal disease.

In addition to the two individuals from family 1505, I also identified one other individual in the Turkish-Cypriot ESRD population who carried the G871C mutation in *COL4A3*. In order to determine the minimal shared haplotype for all these families with the G871C mutation I genotyped one individual from each of the distinct pedigrees that we had access to. These individuals included the 300K SNP data from family 1505, one subject from family 1509 (342), the Greek-Cypriot sample sent by Professor Deltas from one of the original three pedigrees published [173], and this additional individual from the Turkish-Cypriot ESRD population. These samples gave a minimal shared haplotype of 6.05 cM, which corresponds to an estimated MRCA 17 generations ago (95% CI 5-61) (Figure 3.5).

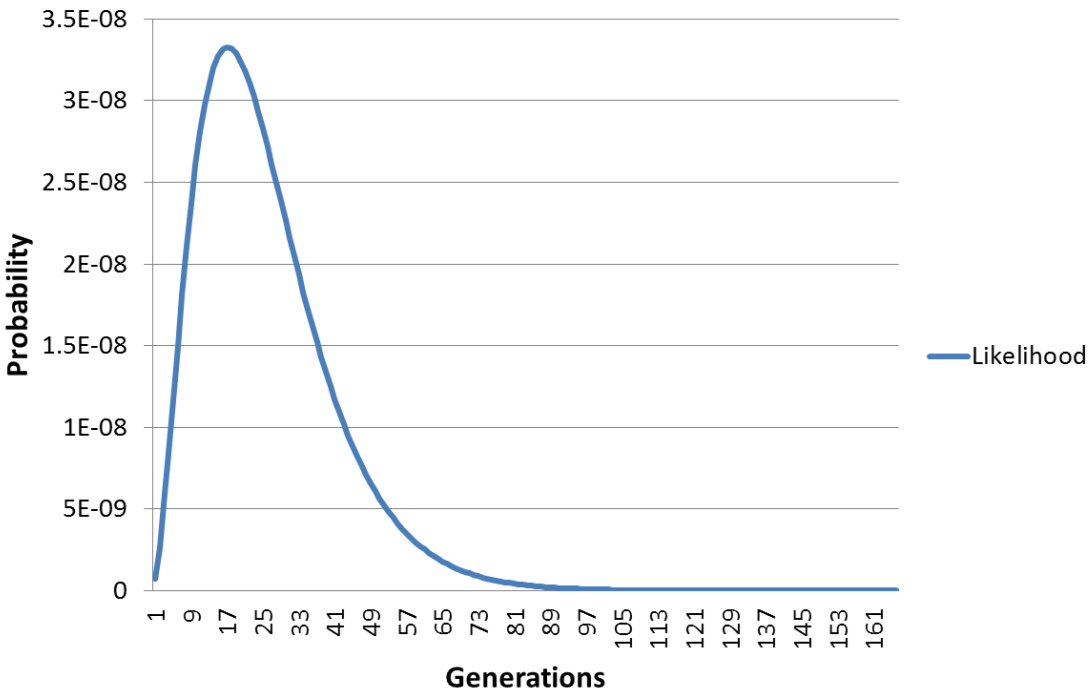


Figure 3.5 Most recent common ancestor (MRCA) estimation for the G871C mutation in *COL4A3*

The MRCA for this mutation is 17 generations using the minimal haplotype shared by individuals from 4 pedigrees not previously known to be related.

Given that all the individuals carrying the G871C mutation identified thus far shared a common haplotype, it is probable that their geographical location within Cyprus would give an idea where the mutation originated (Figure 3.6). I have highlighted those families carrying the G871C mutation in *COL4A3* identified by my research. There is significant geographical clustering of many of these mutations suggesting common founders, notwithstanding migration to the island's capital, Nicosia.

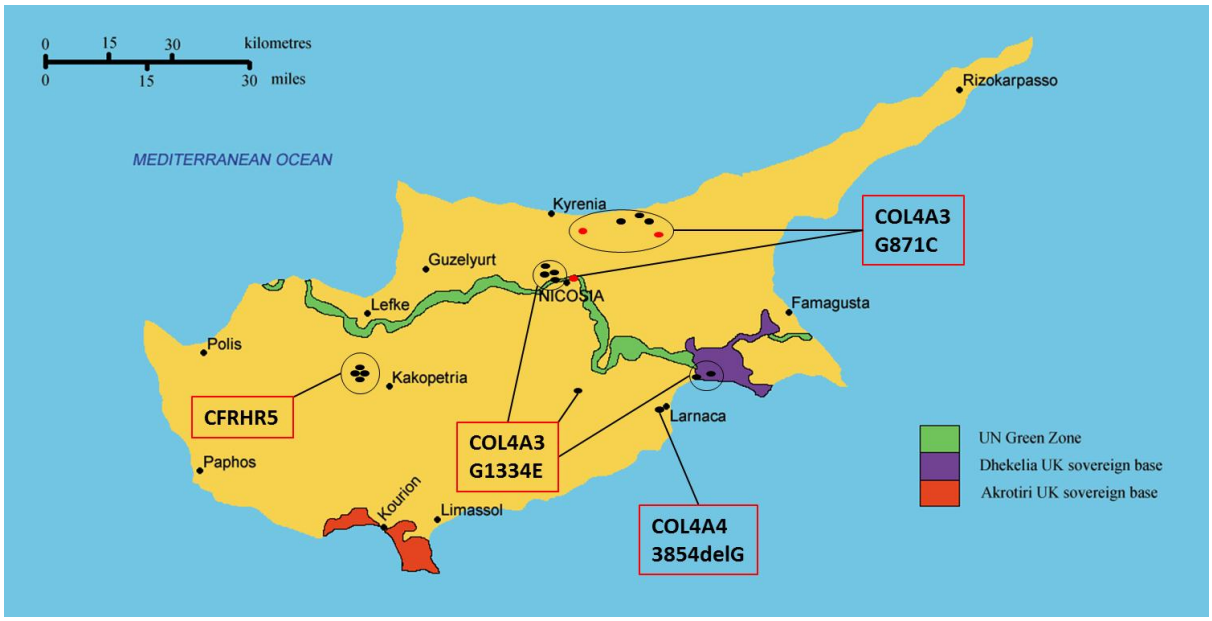


Figure 3.6 Map of Cyprus showing geographical clustering of mutations in *COL4A3*, *COL4A4*, and *CFHR5*

Each oval indicates a distinct family carrying the indicated mutation (adapted from [173]). I have highlighted the three pedigrees carrying the G871C mutation identified by this research in red.

3.3 Interpretation

3.3.01 High incidence rate of renal replacement therapy in Turkish Cypriots

This study presents the first population-based RRT incidence figure from Cyprus, and reveals a higher incidence of RRT than other countries reporting to the ERA-EDTA, with the exception of Turkey [15]. Diabetes is a major cause of ESRD overall and specifically in those under 65 years of age, with rates comparable to those seen in whites in the USA [14]. We found that the high incidence of RRT in Turkish-Cypriots is not due to the specific mutations in *COL4A3, COL4A4,* and *CFHR5* assessed in this study [137, 171]. Finally, a third of Turkish-Cypriot patients start RRT with unknown primary diagnosis. This highlights both the need for earlier detection of these cases and the possibility that there may be other conditions causing ESRD in this population.

3.3.02 Diabetic nephropathy

The Turkish-Cypriot case-mix, with a high incidence of RRT for ESRD due to diabetic nephropathy, is similar to that seen in Turkey and Greece (Table 3.2). Diabetes is also a common cause of ESRD in developing countries around the Eastern Mediterranean, such as Egypt, Kuwait, Lebanon and Saudi Arabia [182]. The incidence of RRT seen in Turkish-Cypriots less than 65 years old is striking (Figure 3.1), and there is evidence that a significant portion of this is attributed to diabetic nephropathy [176]. Currently data concerning the incidence of diabetes in Turkish-Cypriots is lacking. Diabetes is common in Greek-Cypriots and mainland Turkish patients [183-185], but not to the extent seen in some other populations with high levels of diabetic nephropathy, such as the Pima Indians of Arizona [165, 186]. Moreover, childhood levels of obesity in Cyprus are much closer to those seen in other European countries than in the US [187, 188]. It is therefore possible that, in common with other registries, some of these cases reflect a co-occurrence of Type 2 diabetes with ESRD due to an alternative aetiology [189].

3.3.03 Coding 'Uncertain aetiology'

There is a significant group in the Turkish-Cypriot RRT population with unknown diagnosis (69.8 pmp). This group partly reflects late presentation and limited diagnostic investigations. Comparative that data from the Eastern Mediterranean is less robust [179, 189], but there is evidence for a reciprocal relationship with coding for hypertension in many registries [179,

189, 190]. The incidence of hypertensive nephropathy is conspicuously lower in Turkish-Cypriots than in neighbouring countries. In the absence of clear diagnostic criteria for hypertensive nephropathy, it may be appropriate to combine 'hypertension' with 'uncertain aetiology' in *Includes presumed glomerulonephritis not biopsy proven.

Table 3.2, which would imply that there is no established primary renal diagnosis in 40% of patients on RRT in each of the Mediterranean registries studied.

The majority of patients with an unknown diagnosis present clinically with minimal proteinuria (<1 g protein/day) and asymptomatic disease, consistent with a pathological process primarily affecting the renal tubules [179] and similar to the features of medullary cystic kidney disease type 1. A number of families presenting this way in the Greek Cypriot population have shown linkage to the region 1q21, but the gene responsible has not yet been identified [172]. It is possible that the medullary cystic phenotype reflects the final common pathway of a number of genetic and environmental factors that are common in this population.

3.3.04 Alternative explanations for the incidence of RRT in Turkish-Cypriots

Regional variations in RRT incidence may reflect both genetic and environmental factors [9, 165, 166]. Macroeconomic factors that influence regional variations in RRT incidence include *per capita* gross domestic product (GDP) and health-care expenditure [9]. Northern Cyprus has a developing economy, with per-capita GDP that is 76% of that in the Republic of Cyprus, and it is dependent on aid from the Turkish government [191]. RRT has only recently become widely available in TRNC and there is no long-term provision of private dialysis. Economic factors also influence the management of chronic kidney disease (CKD) and co-morbidities, as well as the competing risk of mortality, in the general population [9]. The high incidence of RRT in the Turkish-Cypriot population may therefore reflect suboptimal management of diabetes, hypertension, and associated complications. This has significance for healthcare planning on the island, and underlines the importance of prospective assessment of kidney function in this population.

3.3.05 Incidence rates in the ethnic Turkish population

Table 3.2 and Figure 3.1 show that the incidence of RRT in the Turkish-Cypriot population is similar to that in neighbouring countries in the eastern Mediterranean. These populations

show both genetic [167], and environmental [15] similarities. In order to try to differentiate between these two factors, I also examined the incidence of ESRD in the ethnic Turkish population in Northern Cyprus. This population is hypothesised to share the same environment but have a different genetic make-up. Other useful comparator populations would be that of the Greek-Cypriot population in the Republic of Cyprus and the expatriate Turkish-Cypriot population in the UK or Australia, however that data was not available.

The Turkish immigrant population in northern Cyprus comprises 27.5% of the population in Northern Cyprus [148]. This population is younger than the Turkish-Cypriot population and therefore likely to show a reduced rate of renal disease. Although the same methods for capturing ethnicity and tracking patient data were used as for Turkish Cypriot patients, this dataset was noticeably less complete. Moreover the sample size of the population was very small (76 patients over 8 years).

The incidence of RRT in the Turkish migrant population is not as high as that seen in the Turkish-Cypriot population or in the population of mainland Turkey. The rates are broadly comparable, however, and given the caveats discussed, it is not possible to distinguish further between genetic and environmental factors using this population.

3.3.06 Genetic renal disease in the Turkish-Cypriot population

Congenital factors may also be important in the aetiology of ESRD in this population [192]. Previous estimates of the prevalence of family history of ESRD amongst incident dialysis patients, suggest that 7-15% of Caucasians have a first or second-degree relative with ESRD [17, 43]. This proportion is highest in young adults, non-Caucasians, and for those where ESRD is caused by diabetes or hypertension [17]. We observed a similar rate of familial ESRD in Turkish-Cypriot patients on RRT, which was even higher (37%) in the group with unknown diagnosis, suggesting the existence of conditions that are not yet characterised in this population. For comparison, the rate of RRT for ESRD due to polycystic kidney disease was similar across all registries.

In order to assess the contribution of inherited renal disease we collected DNA from the entire Turkish-Cypriot population on RRT. Previous work has demonstrated a number of important founder mutations and significant geographic clustering for several monogenetic diseases affecting the Greek-Cypriot population of Cyprus, including mutations of *COL4A3*, *COL4A4,*

PKD2, and *MEFV* [81, 171, 174]. Although the *CFHR5* duplication is common in Greek-Cypriots originating from the southern half of the island [193], we did not detect it in our sample of prevalent patients. However, due to the small size of our sample the allelic frequency may still lie within that observed for Greek Cypriots [194]. The high incidence of RRT in Turkish-Cypriots is likely to be due to environmental factors combined with polygenic contributions from many risk alleles.

3.3.07 Geographical evidence for a founder mutation

Two major factors which determine the spread of alleles within a population are the origin and age of a given variant. The geographical clustering of families with distinct mutations enables us to estimate the likely origin of these mutations. CFHR5 nephropathy is believed to have originated within the Trodos mountains of western Cyprus [137]. Previous work has also identified clusters of families with mutations in *COL4A3* and *COL4A3* [173]. The G1334E mutation in *COL4A3* appears to cluster around Nicosia and two other villages further east, while the 3534delG mutation in *COL4A3* has been found only in a village near Larnaka in south-eastern Cyprus.

The G871C mutation in *COL4A3* was described in three families from villages near Kyrenia in northern Cyprus [173]. Although Turkish administered Northern Cyprus (TRNC) extends both to Nicosia and Famagusta, the Kyrenia region in northern Cyprus is most completely situated within that territory. This may explain why the G871C mutation was the only mutation identified by this study. This point of origin is further supported by the identification of this mutation in a family in the UK that can also trace its ancestry back to the Kyrenia district of northern Cyprus.

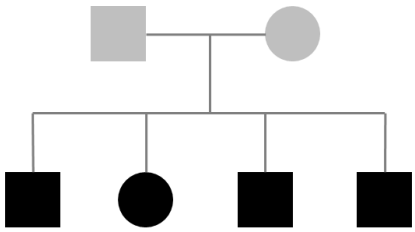
3.3.08 Calculation of the age of the G871C mutation

The existence of a mutation in a population is the result of a number of competing factors. It may exist at steady state levels by recurrent identical mutational events or preserved by heterozygote advantage. Alternatively, it may be on its way to extinction due to negative selection, but still detectable due to recent origin or a founder effect. Allelic age can be estimated from the analysis of the decay over time of linkage disequilibrium between the mutant allele and one or more adjacent marker loci [195]. Alleles inherited from a common

founder will be identical by descent, and will therefore share the same LD between polymorphisms.

Calculation of the age of a mutation using the pattern of linkage disequilibrium at closely linked marker loci can be thought of as using a 'genetic clock' [196]. This relies on the fact that, due to recombination, the extent of LD decays exponentially with generations at a rate that is proportional to the recombination fraction (θ). The true age is the time at which the mutation first occurred. However in many cases, such as with founder effects due to immigration, the aim is to determine to time of the most recent common ancestor (MRCA) of the mutation bearing chromosomes in a population [197]. MRCA calculation was previously performed using micro-satellite markers, however high density SNP arrays can provide a much more accurate haplotype measurement.

We determined the minimal haplotype shared by four discrete pedigrees with the G871C mutation by assessing conflicting homozygosity (CH). CH refers to the situation where individuals are homozygous for different (ie. conflicting) alleles at a given locus (Figure 3.7). The occurrence of CH means that these individuals cannot be identical by descent (IBD) at that locus. Analysis of CH therefore allows the extent of the shared haplotype to be determined.



SNP						
1	Heterozyous	AB	BB	**AA**	AB	Not informative
2	Homozyous – no CH	AA	AA	**AA**	AA	Not informative
3	Heterozyous – no CH	AA	AA	**AB**	AA	Not informative
4	Conflicting homozygosity	AA	AA	**BB**	AA	Informative

Figure 3.7 Conflicting homozygosity (CH) analysis

Each SNP is considered in isolation and CH occurs at SNP 4 in this example. There is therefore a run of three SNPs with no CH, which is consistent with a shared haplotype (identity by descent). The occurrence of CH, indicates that it is not possible for a single haplotype to be shared by all individuals under consideration here.

Unrelated individuals will have short runs of SNPs that appear to be IBD by chance alone. The expected length of such runs varies as a function of the number of individuals compared and the minor allele frequencies of the SNPs in the population from which the subjects are drawn. By contrast, a group of individuals that have inherited a region of the genome from a common ancestor will have much longer runs without CH.

This approach is different to that used by Voskarides *et al.*, who used short tandem repeat (STR) markers to date the G1334E mutation in *COL4A3* [173]. These researchers were not able to determine the size of the G871C haplotype using STR markers as they had only three families with this mutation. I was able to determine the shared haplotype using four families because of the density of markers used in CH analysis. Voskarides *et al* used 22 STR markers over 23 cM to measure a 13.71 cM haplotype, whereas the 300K array used in this investigation has a resolution of <0.2 cM (expected runs of 17 SNPS demonstrating no CH, given a minor allele frequency of 0.22).

The advantage of using CH to determine a shared haplotype is that unlike linkage mapping, it does not require the exact relationship between pedigrees to be known. Moreover, it does not require the determination of marker frequencies in the general population in order to define the boundaries of a shared haplotype. One disadvantage, however, is that this analysis will tend to overestimate the size of the shared haplotype because the actual recombination point is likely to occur before the locus demonstrating CH.

Once the size of the shared haplotype is known it is first necessary to determine the genetic distance between the boundary SNPs and the adjacent informative markers. The probability of no recombination between a marker x that is κ Morgans distant from the mutation in n generations is expressed by the following equation:

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

And the likelihood of recombination between x and the adjacent marker x-1 expressed as:

$$F(x) = T(x-1) – T(x)$$

Assuming T(x-1) = T(x), where x and x-1 are very close to each other, then the probability of recombination between one or both individuals at this point is $F(x)^2 + 2T(x)F(x)$

Consequently, the likelihood that the observed haplotype was transmitted from an ancestor n generations ago is given by combining the likelihood for the contromeric and telomeric shared markers as:

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

The value of n for which L(n) is maximised is therefore an estimate of the number of generations since the MRCA. This equation can be solved iteratively for each additional generation, and the resultant curve plotted (Figure 3.5). This allowed the estimation of the maximal value of L(n) as well as a 95% confidence interval.

Assuming that one generation equates to 25 years, the estimation of an MRCA 17 generations ago means that the G871C mutation originated at least 425 years ago. It is possible, however that the mutation arose earlier, prior to the bifurcation of the different families included in this study [197]. This is similar to the date estimated for the G1334E mutation [173], and is compatible with the theory that most harmful mutations are of relatively recent evolutionary origin [78]. One of the pedigrees used in determining the MRCA for the G871C mutations came from a sample provided by Professor Deltas from the original three pedigrees published [173]. It is possible that the other two pedigrees might define a smaller shared haplotype, in which case the mutation is likely to be older than this calculation.

3.3.09 Phenocopies

As shown in Figure 3.4, family 1505 is a large pedigree that co-segregated microscopic haematuria and mild proteinuria with renal impairment. The pattern of inheritance suggested an autosomal dominant mode of inheritance, and linkage analysis had been performed using SNP genotype data. However this analysis had failed to demonstrate the co-segregation of renal disease with any single genetic locus. Once two individuals were identified with the G871C mutation in *COL4A3* it became clear that linkage analysis had failed in this family due to the presence of multiple phenocopies.

A phenocopy is an individual whose phenotype is identical to the one of another individual whose phenotype is determined by their genotype. Genetic linkage depends on the

association of particular marker loci with a given trait [48]. Since phenocopies exhibit the same trait in association with different loci, erroneous associations will be made in the context of familial linkage analysis. The problem of phenocopies can be only be avoided by increasing the sample size by several orders of magnitude [198], as in the case of genome wide association studies, or minimised by relying on a highly specific phenotype. Characterising a disease phenotype can be difficult, especially when resources are limited. As discussed previously, renal disease is often assessed with surrogate markers, such as serum creatinine, which are influenced by a wide variety of events.

A further question is why there should be so many individuals with renal disease within this family. Given the high incidence of ESRD in the Turkish Cypriot population, it is possible that there are other genetic risk factors for renal disease that segregate independently of the *COL4A3* mutation in this family. Applying the same technique of CH analysis to these phenocopies did not reveal any regions that were IBD, demonstrating that the same genetic factor(s) were not operating in all phenocopies.

3.3.10 Risk of ESRD with the G871C mutation

Familial microscopic haematuria has long been considered a benign condition; however the first description of these Cypriot mutations in Type IV collagen showed a strong association with glomerular scarring and progression to ESRD. Careful follow-up of a large cohort of heterozygous carriers of either autosomal or X-linked mutations in type 4 collagen does suggest that there is a 15% long-term risk of developing renal failure [199]. Moreover, the risk of ESRD in those with heterozygous mutations in *COL4A3* or *COL4A4*, labelled as autosomal dominant Alports, is also around 20%, with an age of onset that is almost always above 40 years [200-202]. Three out of 15 heterozygote carriers are now on RRT in family 1505, giving a risk of ESRD of 20% in this small sample.

Thin basement membrane nephropathy is prevalent in Cyprus. One study showed that 5.45% of all biopsies processed by electron microscopy showed TBMN [203], and up to 53% if only biopsies taken for investigation of isolated glomerular haematuria are considered [26]. An important question is therefore why some individuals with TBMN should progress to ESRD. There is evidence that the R229Q polymorphism in *NPHS2* is associated with proteinuria in Greek Cypriots [134]. However this effect was slight, and there is still a debate whether this variant does indeed act as a disease modifier. Irrespective of whether there are other modifier

genes in the Turkish-Cypriot population, it is probable that these variants giving rise to TBMN are not as benign as originally thought.

3.4 Conclusions and future work

This study is the first complete collection of DNA from an ethnically defined population with ESRD. Renal disease is common in Turkish-Cypriots, and there is a significant rate of familial renal disease. The high incidence of renal disease in this population is not due to the specific mutations in *COL4A3, COL4A4,* and *CFHR5* assessed in this study [137, 171]. The occurrence of the G871C mutation in this population is evidence for gene flow between Greek-Cypriot and Turkish-Cypriot. Given the size of the population sampled it is not possible to say that the other mutations do not occur in Northern Cyprus.

One third of Turkish-Cypriot patients start RRT with unknown primary diagnosis. This proportion may be higher if we accept the limitations of the diabetic and hypertensive diagnostic categories. This highlights the need for earlier detection and more detailed investigation of renal disease in this population. It is tempting to speculate that there may be other conditions causing ESRD in this population. Certainly there are both genetic and environmental differences between the Turkish-Cypriot population and the better studied populations of Western Europe and North America that may give rise to novel renal pathology.

Diabetes is a major cause of ESRD overall and specifically in those under 65 years old, with rates comparable to those seen in whites in the USA. In the Afro-Caribbean population of the southern United States it is notable that the highest rates of familial renal disease are found in those with diabetic and hypertensive nephropathy [17], both diagnostic categories that are common in the Turkish-Cypriot population. It is possible that the familial clustering observed with these conditions is due to multiple additive risk factors, rather than few variants with large effect size. It is also likely that environmental factors are important.

I have shown that the incidence of RRT in the Turkish-Cypriot population of Northern Cyprus is amongst the highest in Europe. It is uncertain what is driving the high incidence of ESRD around the eastern Mediterranean. Genetic association studies might help our understanding at a population level, particularly if contrasted with populations exhibiting a low incidence of renal disease, such as that of Iceland [204]. This population therefore provides an opportunity to look for genetic factors associated with an increased risk of ESRD in Cyprus.

Chapter 4 Familial kidney disease

4.1 Introduction

Analysis of the Turkish Cypriot population has shown that incidence of ESRD is amongst the highest in Europe [176]. Some of this increased incidence may be due to genetic factors, but despite close similarities with the Greek Cypriot population, we did not find that this was due to four known mutations that are associated with renal disease in that population [137, 171]. It is therefore likely that other, as yet unidentified, genetic risk factors that operate within the Turkish Cypriot population give rise to the high rates of renal disease observed.

Analysis of the Turkish Cypriot ESRD population showed a high proportion of patients with a family history of renal disease. This was especially true for those with an uncertain aetiology, who often present clinically with a minimal proteinuria (<1 g protein/day) and asymptomatic disease, consistent with a pathological process primarily affecting the renal tubules. Given the limitations of diagnostic work-up in some countries of the Middle East, Neild *et al* have proposed a revised coding system that essentially divides individuals into the broad categories of glomerular or tubular phenotype. The glomerular phenotype is a clinical history consistent with glomerulonephritis, but a renal biopsy may not have been performed [179]. Clinically this means a presentation with macroscopic haematuria, nephritic syndrome, fluid retention, malignant hypertension, or via routine investigations. By contrast, the tubular phenotype includes patients who reach ESRD with minimal symptoms, minimal proteinuria, no peripheral oedema, and without malignant hypertension [179].

In order to try to identify additional genetic factors operating in the Cypriot population, I studied a large Cypriot family with an inherited form of renal disease. This family originates from the south of the island, and a number of individuals have since emigrated to south London. The clinical phenotype in this family is of proteinuria, microscopic haematuria, and renal impairment that appear to be inherited in an autosomal dominant fashion. Five individuals have had a renal biopsy, but there was no common pathological diagnosis. Four individuals have received a renal transplant, one of which has survived for 16 years without any evidence of recurrent disease.

This family came to our attention after the identification of *CFHR5* nephropathy [137]. The family had heard about *CFHR5* nephropathy through the media and wanted to be tested for this condition, on account of their extensive family history of renal disease and ethnicity. DNA had been collected in 1994 from some of the family by researchers at the University of Cyprus.

This had been used for linkage analysis, employing microsatellite markers, but no clear linkage had been identified. Pedigree information had also been collected from some of the individuals living in London by their nephrologist at St Georges' Hospital; however no analysis had been performed on these samples.

4.2 Results

The pedigree of family 1508 is shown in Figure 4.1. The first stage of my analysis in this pedigree was to show that none of these individuals carried the duplication in *CFHR5* nephropathy or the mutations in *COL4A43* and *COL4A4* already described [137, 171] (data not shown). It became apparent that it would be difficult to define a precise phenotype as renal disease in many individuals was mild and non-specific. The condition also shows incomplete penetrance as one individual, 1508 202, is an obligate carrier and yet has survived into her 90s, with no evidence of renal impairment on laboratory investigation.

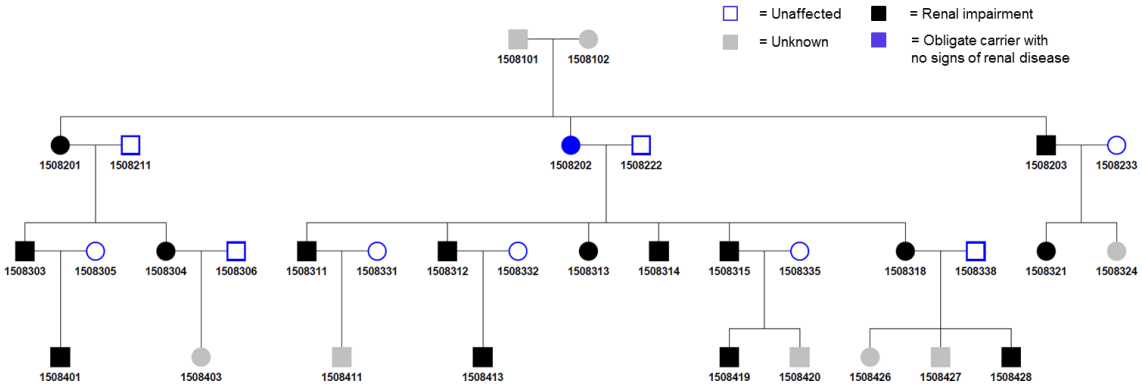


Figure 4.1 Pedigree of family 1508

Affected individuals include all those with evidence of renal impairment. The individual in blue (1508 202) is an obligate carrier with no phenotypic evidence of renal disease. Unrelated spouses are depicted as unaffected. The remaining individuals are phenotypically normal but may go on to develop renal impairment and thus were not included in the analysis.

There were just three individuals in the UK with strong evidence of renal disease. The first of these (1508 304) reached ESRD aged 34 and was transplanted. The second (1508 321) was under follow-up in the midlands for stable renal impairment (CKD stage 3). She was referred in 2009 with a creatinine of 115 umol/l (eGFR 44 ml/min), intermittent haematuria, and variable proteinuria (albumin creatinine ratio (ACR) 42.5 – 184.3), in the context of a new diagnosis of type II diabetes, hypertension, and mild obesity. The third (1508 401) had had a renal biopsy aged 16 for microscopic haematuria and proteinuria. After 18 years of follow up he had persistent urinary abnormalities with mildly impaired renal function (creatinine 120 umol/l, eGFR 64 ml/min).

In order to identify the mutation responsible for the renal disease in this pedigree I collected DNA samples from 33 individuals. I was able to access the renal biopsy reports from four individuals; the conclusions from which are summarised in Table 4.1. DNA samples from 25 individuals were used to generate 300k SNP genotypes.

ID	Date	Age	Locn.	Findings
304	1964	17	UK	5 glomeruli; GBM increased refractility Tubules well preserved with no evidence of vascular disease or interstitial inflammation Non-diagnostic: possibly early membranous
304	1974	27	UK	Many glomeruli hyalinised, others segmental sclerosis with moderately increased mesangial cellularity Severe tubular atrophy and interstitial fibrosis Moderate interstitial inflammatory infiltrate Advanced nephrosclerosis
314	1989	37	UK	'Alport Syndrome' - report not available No haematuria or hearing loss
313	1995	46	Cyprus	Endothelial and mesangial cells normal Vacuolated epithelial cells with focal podocyte fusion EM: no dense deposits; normal GBM thickness with focal thinning and thickening but no splitting
315	1995	35	Cyprus	2 glomeruli, no increase in mesangial matrix Epithelial cells prominent with many micorvilli EM: Lamina densa quite uniform, no dense deposits No evidence of Alports
401	1995	18	UK	Cortex & medulla; 10 Glomeruli, none sclerosed; IF negative, few epithelial foot process effacement Minor tubular atrophy and ischemic wrinkling EM: no dense deposits or lamina densa splitting/duplication Non-diagnostic

Table 4.1 Renal biopsy findings in family 1508

Genotypes were analysed using PedCheck, which demonstrated that two of the Cypriot samples gave rise to a large number of major Mendelian errors. Individual 428 and his father gave rise to 1316 Mendelian errors, while 321 and her mother gave rise to 666 Mendelian errors. This means that the samples were either from the wrong individuals, or that the parentage of the individuals was not as we were led to believe from the pedigree. Paternal discrepancy is common, affecting a mean of 3.7% of the population [205], and may have contributed to the failure of the microsatellite linkage analysis performed in Cyprus. For this reason, those individuals were excluded from subsequent linkage analysis.

Preliminary linkage analysis was performed on this family using a subset of 9600 SNPs in order to reduce the analysis time (Figure 4.2). Genotypes were re-analysed, after the exclusion of 428 and 321, and 17 Mendelian errors were detected. These were thought to correspond to miscalled SNPs on the chip and were therefore discarded. The remaining genotypes were analysed using GENEHUNTER [154] and SimWalk, using the EasyLINKAGE interface [153]. The genotypes were divided into sets of markers (80, 100, or 120) on each chromosome and analysis was performed in all individuals in each set, assuming a dominant model of inheritance.

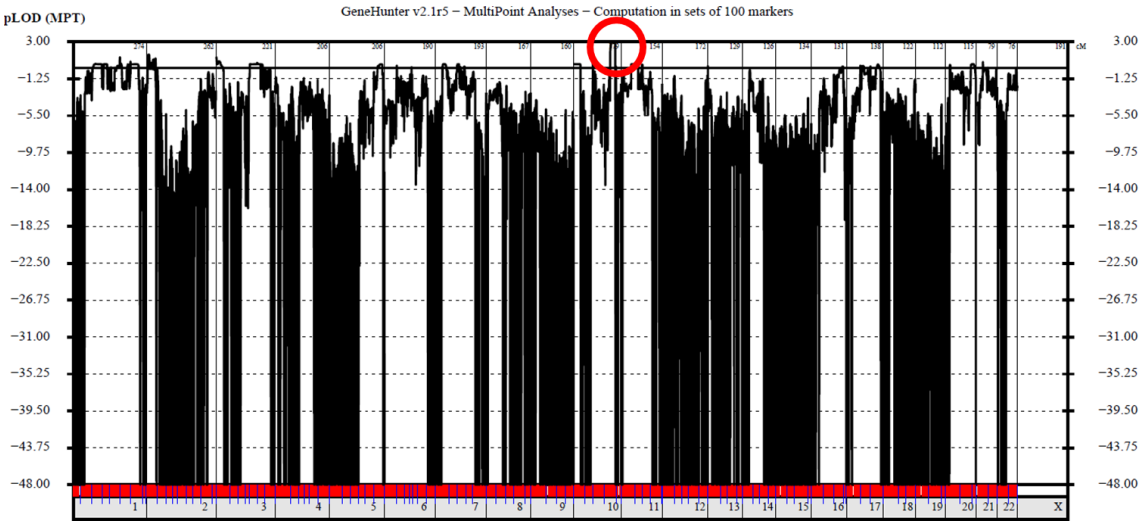


Figure 4.2 Linkage analysis of family 1508

Genome-wide linkage analyses of this pedigree using GENEHUNTER yielded a maximal LOD score of 2.97 in association with a region on Chromosome 10 (circled). Analysis of this peak on chromosome 10 with SimWalk gave a maximal LOD score of 4.2.

The maximal LOD score using GENEHUNTER to analyse this pedigree was 2.976 in association with a region on chromosome 10 (Figure 4.2). Using the maximum probability haplotype reconstruction in GENEHUNTER I showed that the haplotype in this region was shared by all affected individuals (Figure 4.3), suggesting that individuals had been excluded by the program from analysis. Reanalysis of the linkage peak on chromosome 10 with SimWalk gave a maximal LOD score of 4.187. The region on chromosome 10 extended 15 cM and did not include any genes known to cause renal impairment.

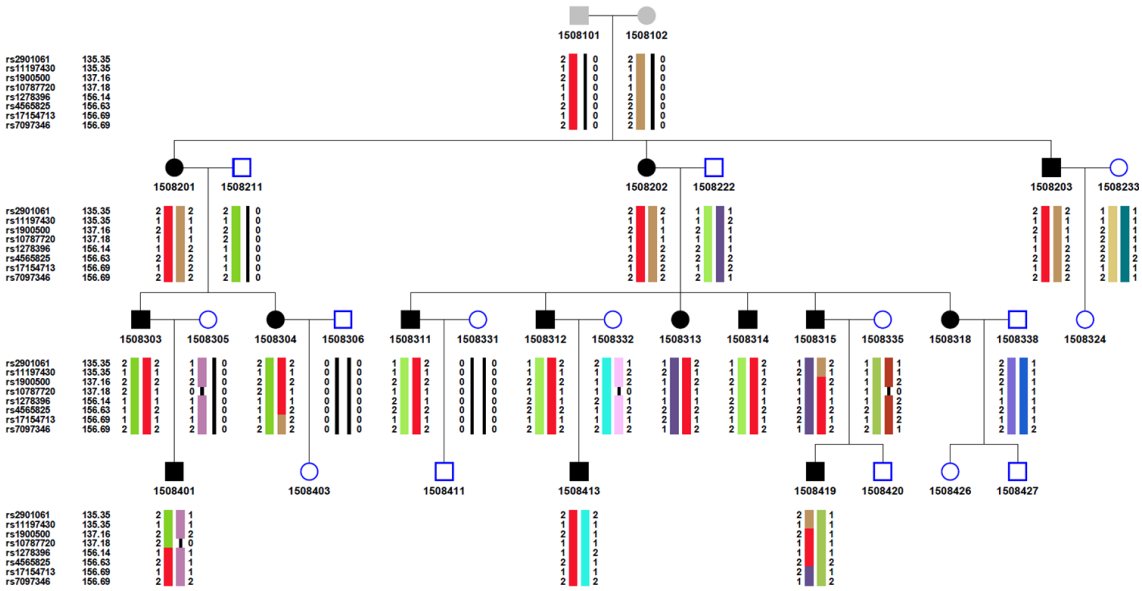


Figure 4.3 Haplotype reconstruction of the linked interval on chromosome 10

Eight SNPs are shown indicating that informative meiosis have occurred to limit the shared haplotype, shown in red, in individuals 304, 315, 401, and 419.

Four individuals in this pedigree had progressed to ESRD (Figure 4.4), all of whom shared the linked region on chromosome 10 (Figure 4.3). The existence of a shared haplotype was also confirmed with conflicting homozygosity analysis (Figure 4.5). CH analysis was used to determine which regions of the genome were identical by descent in all combinations of 11 affected family members that included the four individuals who had proceeded to ESRD. As illustrated in Figure 4.5, the addition of two additional individuals to two first cousins once removed significantly reduces the regions that are IBD to just three regions.

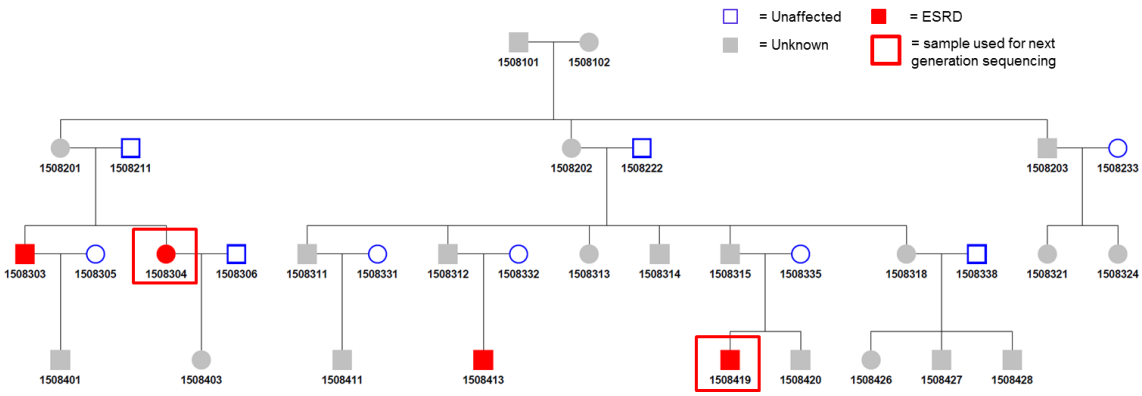


Figure 4.4 Pedigree of family 1508 showing those individuals who had reached ESRD

The individuals used for next generation sequencing are indicated by red boxes.

In order to identify the causative variant from within this linked region, two individuals were chosen for whole genome sequencing. I selected two individuals in order to be able to filter out variants that were not shared between them, rather than relying solely on the linkage data. Both individuals had ESRD, so that we could be most sure they were not phenocopies. Lastly these two individuals had the maximal genetic distance between them, shown in Figure 4.4. These individuals were first cousins once removed, and were separated by five meioses. Consequently they would share, on average, 3.125% of their genetic sequence. Table 4.2 shows the depth and coverage of the sequencing achieved. We identified ~4 million variants.

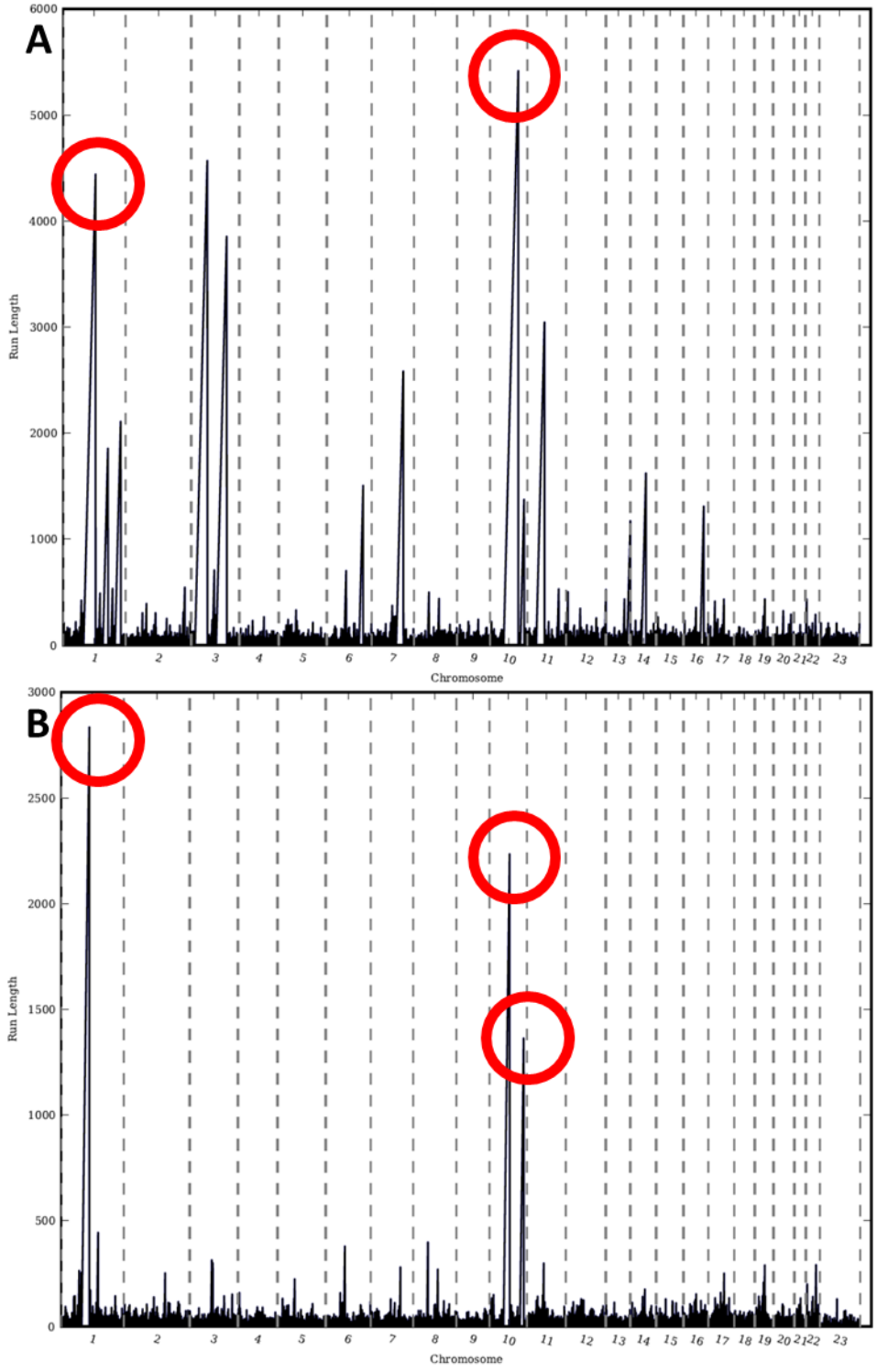


Figure 4.5 Conflicting homozygosity (CH) analysis in family 1508

CH analysis of the two individuals used for next generation sequencing (A) and of all four individuals with ESRD (B). Run length above 300 SNPs indicates a significant shared haplotype. The shared regions circled in B have also been circled in in A, and were used to define the shared region that was used to filter out variants shared by all four individuals with ESRD in Table 4.3.

In order to determine which of the variants identified by WGS was responsible for renal disease in this family I employed a number of filtering steps, based on assumptions that parallel the reasoning behind the linkage analysis in this family. First, the sequencing data was examined to see whether either individual harboured any single mutation that would have been sufficient on its own to cause ESRD. This was done by searching the data with a list of around 50 genes that have well known associations with renal disease, such as collagen, type-IV, alpha-5 (*COL4A5)* or *NPHS1* [23]. While both individuals were carriers for a number of polymorphisms in these genes, none were present in the database as pathogenic. In essence this answers the question of whether either of the two individuals is a phenocopy; that is, whether they have ESRD on account of a known genetic cause.

Category	**Metric**	**1508 304**	**1508 419**
Miscellaneous	Gender	Female	Male
Genome coverage	Fully called genome fraction	0.973	0.971
Exome coverage	Fully called exome fraction	0.975	0.977
Genome coverage	Gross mapping yield (Gb)	197.377	197.135
Genome variations	Heterozygous SNP count	2098591	1996406
Genome variations	SNP transitions/transversions ratio	2.133	2.138

Table 4.2 Sequencing metrics from whole genome sequencing data

Subsequent filtering steps were used to further reduce the number of variants identified by whole genome sequencing in this study (Table 4.3). Given that this disease appeared to be an autosomal dominant condition, I looked at only heterozygous variants that were shared between the two individuals. I hypothesised that the causative variant would result in a change to the amino acid sequence of a protein. Lastly I assumed that this variant was novel, and therefore not present in the databases with a MAF > 0.005. Four million heterozygous SNPs were thereby reduced to six novel variants that were taken forward for further analysis. The pathogenicity of these six novel variants was assessed by *in silico* analysis (Table 4.4). Four complementary programmes were used to investigate whether the variants identified showed evolutionary conservation across vertebrate species and whether the predicted amino-acid alteration would affect the function of a particular protein domain.

Filtering step	**Number of variants**
Total heterozygous SNP count in each individual	2098591 and 1996406
Shared variants	78023
Coding sequence and splice site	911
Component: Deletion/Insertion/SNP/Substitution	431
Impact: All except no change and synonymous	217
Shared by all four individuals with ESRD	69
Novel	6

Table 4.3 Filtering of heterozygous SNPs identified by whole genome sequencing

Evaluation of the six variants identified by WGS did not reveal any that were clearly deleterious (Table 4.4). The poor correlation between scores is indicative of the limitations of using these tools. The variant in *PCDH15* scored most highly of the six tested with MutationTaster and PolyPhen2, suggesting that it was most likely to be pathogenic. It also scored lowest with SIFT, suggesting that it would not be tolerated, although the score was above the threshold of 0.05. By contrast the variants in *GRK5*, *DMBT1* and *DHX32* all had very low scores suggesting that they were benign changes. Evolutionary conservation was assessed using PhyloP to compare conservation among 44 vertebrate species. The positive PhyloP scores for *PCDH15*, *GRK5*, and *DHX32* show greater conservation at that individual nucleotide than would be predicted by chance.

Gene	**Nucleotide substitution**	**Amino Acid substitution**	**Mutation taster**	**SIFT score**	**PolyPhen2 score**	**PhyloP conservation**
PCDH15	G5982C	S1862C	3.05	0.11	0.871	1.41583
GRK5	G1143A	R303H	0.79	0.99	0.068	1.28574
INPP5F	C3140G	T991S	1.58	0.25	0.000	0.197701
DMBT1	CAC3145AGT	H1103S	2.43	0.94	0.001	0.193882
METTL10	C58T	G7S	1.53	0.23	0.034	-0.70011
DHX32	C2630T	V713M	0.57	0.93	0.001	1.62439

Table 4.4 *In silico* predictions for the six novel variants identified by whole genome sequencing

In parallel to the *in silico* analysis I designed tetra-primer ARMS PCR reactions to genotype the Turkish Cypriot ESRD population for the six novel variants generated by the analysis of the whole genome sequencing data from this family. Previous filtering had defined a novel variant as having a minor allele frequency of < 0.005 in the dbSNP database. However, I hypothesised that data from relatively few Cypriots had been included in dbSNP and therefore that some of these variants might be more common on the island, and therefore not causative. Three of these variants have significant minor allele frequencies, and were therefore not investigated further.

Gene	Turkish Cypriot			Greek Cypriot		MAF
	Subjects tested	Het	Hom	Subjects tested	Mutation positive	
PCDH15	270	0	0	146	0	0.000
GRK5	12	4	0	43	4	0.073
INPP5F	279	0	0	172	0	0.000
DMBT1	92	13	0	46	6	0.068
METTL10	87	17	1			0.094
DHX32	186	0	0	96	0	0.000

Table 4.5 Frequency of novel variants in the Turkish-Cypriot population

The same tetra-primer ARMS PCR reactions were used by our colleagues in Nicosia to genotype additional Greek-Cypriot patients with renal disease. This gave us additional information on the frequency of these alleles in the Cypriot population, as well as the potential to detect additional individuals with the same condition. No individuals were found to carry any of the variants described in *PCDH15*, *INPP5F*, or *DHX32*.

In order to investigate the function of the remaining three variants identified by next generation sequencing in this family, I obtained a culture of primary urothelial cells from four individuals in this family (Figure 4.6). These patient derived cells were genotyped using tetra-primer ARMS PCR on cDNA isolated from cultured cells (Figure 4.6). Analysis of cDNA not only confirms the genotype, but it also demonstrates that the variant is expressed as mRNA. Of note, one individual had a functioning renal transplant at the time of urothelial cell collection, however all the clones successfully isolated from this individual were derived from the transplanted urothelium (patient C3 in Figure 4.6).

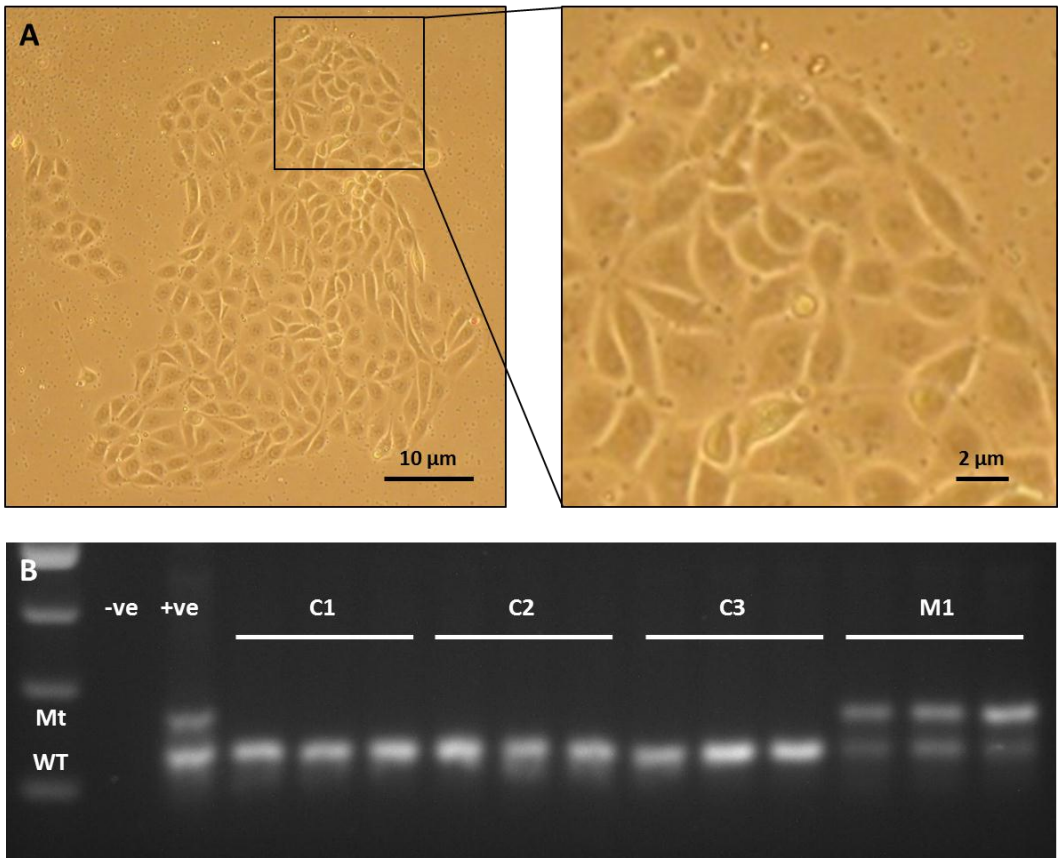


Figure 4.6 Culture of patient-derived urothelial cells in vitro.

A. Typical appearance of a type I colony showing the smooth-edged contours and cobblestone-like cell morphologies. B. Genotyping was performed using tetra primer PCR for the T991S variant in *INPP5F* on cDNA extracted from cultured urothelial cells. Genomic DNA was used as a positive control (+ve) for the presence of the mutant PCR product (Mt).

Gene expression was analysed in urothelial cells collected from individuals in this family in the UK (Figure 4.7). I analysed three independent colonies from two control patients (C1-2) in the family and two individuals affected with renal disease (C3 and M1). I used real-time PCR to examine the expression of a number of genes that have been associated with the urothelial lineage in order to confirm the probable origin of these cells. I then looked at the expression of the three candidate genes from the sequencing experiment. There was no effect of genotype on the expression of these genes in this family.

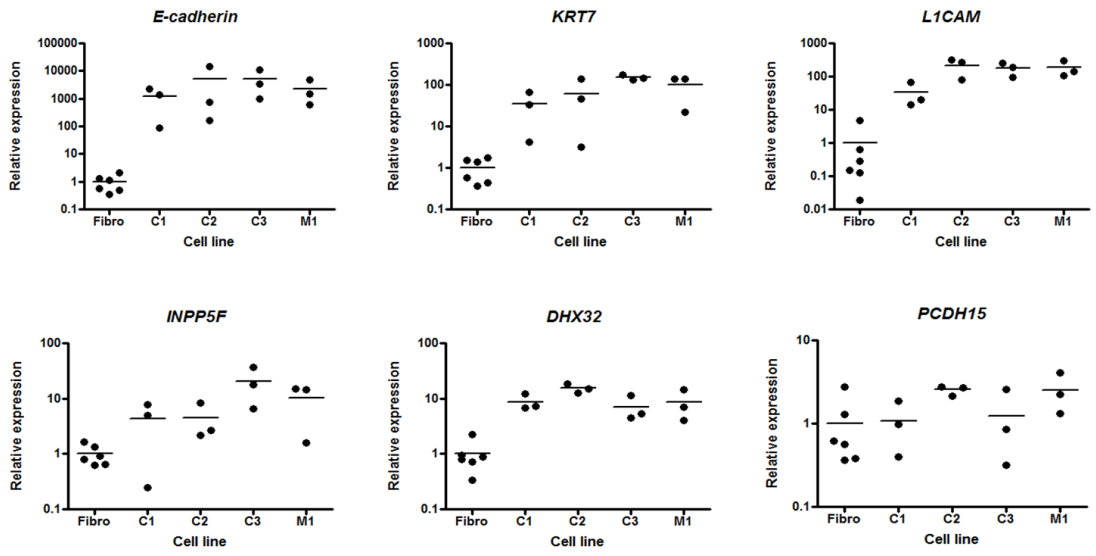


Figure 4.7 Gene expression in patient-derived urothelial cells

Expression of *E-cadherin (CDH1)*, *keratin 7 (KRT7*), and L1 cell adhesion molecule (*L1CAM*) was greater in all urothelial clones compared to a reference population of primary fibroblasts. There was no effect of genotype on the expression of the genes containing novel variant in this family (*INPP5F, DHX32,* or *PCDH15*).

4.3 Interpretation

4.3.01 Linkage analysis

In order to investigate the genetic basis of renal disease in this family I used linkage analysis to determine whether a single genetic region co-segregated with disease. I generated different LOD scores using different analysis programs; however haplotype prediction and CH analysis confirmed that a region on chromosome 10 did segregate with renal disease. The different result from the two programs used highlights a number of limitations with traditional linkage analysis.

Linkage is usually reported as a logarithm of the odds (LOD) score. This score compares the likelihood of obtaining the test data if the two loci are indeed linked, to the likelihood of observing the same data purely by chance. At a given linked region, each additional recombination (i.e. the addition of another affected first-degree relative to the pedigree) will add a maximum of $\log_{10}(2) = 0.3$ to the LOD score. Consequently the maximal LOD score generated by GENEHUNTER (= 2.97) is below that which would be expected from 14 affected individuals (14 x 0.3 = 4.2). This indicates that either some individuals do not share the disease haplotype, or that some individuals have been excluded from analysis for computational reasons due to the size of the pedigree.

GeneHunter uses the Lander-Green algorithm to generate an exact solution to the linkage analysis in question [154]. As such it is limited to a pedigree size of around 20 people (the pedigree in Figure 4.1 contains 33 individuals). By contrast SimWalk uses a Markov chain Monte Carlo algorithm which allows it to consider the underlying possible inheritance data in proportion to their likelihood [155]. This generates an estimate rather than an exact result, but it enables the program to process pedigrees that are an order of magnitude larger. Given the higher score seen with SimWalk and the results of CH analysis, it is clear that this low score is a limitation with the GENEHUNTER program.

The significance of the LOD score depends both on the number of individuals included, and also on the accuracy of the disease phenotype. Unfortunately the early stages of chronic kidney disease are common and non-specific. Chronic kidney disease (CKD) may affect up to 15% of the total population [206, 207], and microscopic haematuria may be found in a similar proportion [208, 209]. In the absence of additional markers of syndromic renal disease, it is

therefore impossible to say whether the abnormal findings in a given individual are due to an early or mild form of the disease in question or due merely coincidence.

Parametric linkage analysis also assumes that we know the mode of disease inheritance. The family studied here is not in-bred, and the pedigree is consistent with a highly penetrant, autosomal dominant condition. However it is possible that there is more than one gene involved, particularly as renal disease is relatively common in this population, as discussed in the previous chapter [176]. This might explain why one individual has no phenotypic evidence of renal disease, and is therefore an obligate carrier.

Another limitation with the traditional linkage approach concerns the size of the linked region. The 15 cM linked interval contained nearly 170 genes, none of which appeared to be directly related to renal pathophysiology. High density SNP arrays such as that used in this project give high resolution of the boundaries of a given linked region. However, the occurrence of homologous recombination in meiosis occurs at a far lower frequency than the information that 300,000 markers can provide. In contrast to the determination of the LOD score, determination of the linked region is a stochastic process that depends on the recombination frequency at a given locus. Thus while recruiting additional individuals to the study can give additional power to the LOD score, it does not necessarily reduce the size of a linked region. Even very large pedigrees may not give a clear association with just a single gene, as seen with MCKD1 [210].

4.3.02 CH analysis

As discussed in the previous chapter, a major limitation with linkage analysis is the existence of phenocopies. In family 1505 discussed in the previous chapter, 4 of 19 affected individuals were found to be phenocopies. This rate would reduce the strength of linkage association significantly. I therefore also utilised CH analysis in family 1508 to look for regions that were identical by descent (Figure 4.5). Given the existence of 25% phenocopies in the 1505 pedigree, we chose to look for shared regions in multiple subsets of the pedigree.

We arbitrarily selected to investigate all combinations of 11 (out of a possible 14) affected individuals that included the four subjects with ESRD. This approach is equivalent to performing parametric linkage analysis on combinations of 11 affected individuals to identify all possible linked regions. However, because CH analysis is so much faster it was possible to

perform this sequential analysis in a few minutes, rather than the days it would have taken with linkage analysis.

I assumed in this analysis that none of the individuals with ESRD were phenocopies. Although urinary abnormalities and elevated serum creatinine may be non-specific, ESRD is still a rare diagnosis. Moreover, all of these individuals reached ESRD at a young age, further increasing the specificity of this diagnosis. By fixing on combinations involving the four subjects with ESRD this analysis looked to see which individuals shared regions that were IBD with them. It was then possible to evaluate the strength of evidence for renal disease in these individuals. The regions that are IBD between these four individuals are shown in Figure 4.5. I did not detect any linked regions in this subset analysis that corresponded to genomic regions containing genes that are known to be linked to renal disease.

4.3.03 Whole Genome Sequencing

Genotyping technologies have enabled the routine assessment of common variants at up to a million sites across the genome in thousands of individuals and have increased our understanding of human genetic diversity and its biological and medical impact [211, 212]. Whole genome sequencing costs have fallen from the >$100 million cost of the first human genomes to the point where individual labs have generated genome sequences in a matter of months for a material costs of only a few thousand dollars [70, 213, 214]. Sequencing technologies use a variety of genomic microarray construction methodologies and sequencing chemistries [71].

Whilst the cost of the various technologies available is the determining factor for many, it is also important to consider the individual methodology. We utilised human genome sequencing using unchained base reads on self-assembling DNA nanoarrays [211]. In common with other technologies the first stage of this technology is the construction of a library of 400 nucleotide genomic DNA fragments. These fragments are then circularised and a series of directional adaptors inserted following recursive cutting with type IIS restriction enzymes. The resulting circles are then amplified to form hundreds of tandem copies in palindrome-promoted coils, referred to as DNA nanoballs (DNBs). The DNBs are then bound to patterned silicon arrays, where they can be read using high accuracy combinatorial probe anchor ligation. This generated independent reads up to 10 bases adjacent to each of eight anchor sites, resulting in a total of 31- to 35-base mate-paired reads (62-70 bases per DNB).

The sequencing data that was returned to us contained information at four levels of analysis. First, the vast majority of variants were composed of single nucleotide substitutions, small insertions or deletions. These were defined as variants located within one base of each other, with the result that they would affect the same amino acid codon. The remainder of the analysis was directed at other, less frequent, genetic variants, namely: copy number variation, chromosomal translocations, and mobile element insertion. These are harder to detect, as they depend on the reliable detection of chromosomal breakpoints, which in turn depends on the read depth, and require *de novo* sequence assembly.

Quality metrics for the data generated show that our results were in keeping with other sequencing technologies. Over 97% of both genomes and exomes were fully called. As shown in Table 4.2, there were around 2 million of these variants per genome, which is keeping with the expected number of this kind of data. Moreover, the SNP transitions/transversions ratio was 2.133 and 2.138, which is again close to the expected ratio of 2 across the whole genome and slightly higher in exomic regions.

4.3.04 Variant filtering

In order to reduce the number of possible variants I filtered the WGS data in a number of ways. First, I assumed that renal disease is caused by a single variant in a gene, which is inherited in an autosomal dominant fashion in this out-bred pedigree. This variant would therefore be heterozygous in both individuals that were sequenced. Consequently, the first filtering step was to search for all those variants which occurred only once in each subject. The number of variants identified at this stage is related to the genetic distance between the two individuals we chose to sequence. Given that they are separated by five meioses they ought to share no more than $1/2^5$ = 3.125% of the genome. We observed that 3.811% of the mean 2047498 variants were shared, which supports this interpretation.

The definition of a heterozygous variant is important, as it relates to the quality of the sequencing read. Each variation from the reference genome is scored for quality. This variation score reflects both the depth of read as well as the number of minor allele reads. Where the ratio of minor allele to reference reads is in excess of 20/40, this is scored with high confidence. Where the ratio is above 10/20, there is only a 10% chance of the minor allele read being correct, and thus the score is low. Below 10/20 no call is made. This impacts on how we filtered for common heterozygous variants. The highest confidence can be attached

to variants called as 1/0 in each individual; however we chose to include those called as 1/N (where 0 is reference, 1 alternative, and N uncalled).

Next, I elected to concentrate on just those variants which involved the coding sequence and splice sites of known proteins. This significantly reduced the number of variants by a further 98.832%, which is in broad agreement with the estimate that the total human exome comprises 1.5% of the genome [215]. The exome is believed to harbour much of the phenotypic variability in an individual, and in contrast to association studies the vast majority of genetic variants with a major phenotypic impact have been discovered within protein coding regions [216]. The number of variants was then reduced by a further 76% by discarding all those which did not change the amino acid sequence of a protein. This left us with a total of 217 single nucleotide polymorphisms, insertions, or deletions shared between two first cousins once removed.

In order to reduce the 217 variants even further, I then used the results of the previous linkage analysis. By looking at just those regions which are IBD in all four individuals with ESRD (Figure 4.5B), encompassing two regions on chromosome 10 and one on chromosome 1, the number of variants of interest was reduced further to 69. It remains possible that one or more of these individuals is a phenocopy; however, I thought this was a necessary assumption to make in order to reduce the number of variants to a manageable level.

With so few variants remaining it became possible to study them in more detail. In particular we wanted to know what proportion of these were novel; the assumption being that if one of these variants did cause ESRD it would not be common in the population. We therefore determined the minor allele frequency for each variant by looking at the NCBI and 1000 genomes databases [217, 218]. A problem with this approach is that these datasets do not include any phenotypic information, and so it is possible that variants with reduced or incomplete penetrance could be included unknowingly. Early estimates that ~80% of an individual's nonsynonymous variants are common in the population [215] look to be increasingly conservative as more and more sequence data is accumulated, making this possibility more likely. Consequently all variants with MAF > 0.005 were excluded.

4.3.05 In silico prediction of pathogenicity

This left us with a list of six novel variants to take forward to more detailed analysis. As indicated previously, none of the genes in the linked intervals appeared to be directly related to renal pathophysiology. One approach was to use bioinformatic tools to look *in silico* for the consequences of these polymorphisms. A second approach was to look directly at the frequency of these variants in the Cypriot population. This could tell us both whether the variant was common, and also if might be associated with renal disease in another family. Lastly we could study any promising candidates *in vitro*, using both cultured cell lines and primary cells derived from patients in this family.

One of the consequences of the next-generation sequencing technology is the generation of vast numbers of genetic variants in an individual, many of which have to be tested for their disease-causing potential. Automated pre-evaluation of sequence variations can help to direct the subsequent in-depth analysis to the most promising candidates [158, 219]. The evaluation of variants is based on two main factors: the evolutionary conservation of a particular sequence across species and the predicted effect of an amino-acid exchange on the function of a particular protein domain in a related protein [159]. Evaluation tools are typically trained on a dataset consisting of known pathogenic mutations and compared to a dataset of benign SNPs [158, 160].

4.3.06 Population genotyping

Most of the genetic variants in any one individual have been previously observed in other individuals. However, the proportion of previously discovered SNPs and novel SNPs is changing over time as databases are updated. Given the small size of the Cypriot population it is likely that there are genetic variants in this population that have not yet been deposited in public databases, but which are still common on the island [218].

We therefore designed a tetra-primer ARMS PCR reaction to study each of these variants in a population that we had access to: the Turkish Cypriot ESRD population (Table 4.5). However this approach does confuse two important issues. First, by looking at a population with ESRD we are potentially selecting for those variants which are associated with renal disease. Second, it would have been better to look for these variants in a representative sample of the Greek Cypriot population, since that is the population from which the family comes. As

discussed earlier, however, there is evidence of gene flow between the populations. The differences that do exist between the populations make the discovery of these variants if anything more significant.

The result of this experiment shows that three of these variants were associated with significant minor allele frequencies. We were able to interrogate the primary renal diagnosis in each of the individuals carrying these variants, and it was apparent from some diagnoses that these variants could not be the sole cause for renal failure. Consequently there were now only three variants to consider for function assessment *in vitro*.

4.3.07 Candidate gene selection

In the absence of any clear *in silico* prediction that one of these three variants was most likely to be pathogenic, we had to prioritise the best candidate for functional studies. This was done by performing a thorough literature search on each of these candidates, as well as looking for evidence of gene expression in the kidney, and if possible, evidence of a renal phenotype from animal knockout studies.

Protocadherin-15 (*PCDH15*) has been implicated in the development of cochlear hair cell stereocilia. Mutations in *PCDH15* cause Usher syndrome, which is characterised by congenital deafness, vestibular dysfunction and retinitis pigmentosa [220]. Deletion in the mouse produces the Ames waltzer phenotype, while a rat *Pcdh15* knockout also shows severe defects in cochlear hair cell stereocilia, collapse of the organ of Corti, and marked reduction of ganglion cells. There is no reported renal phenotype.

Although *PCDH15* lies within the shared haplotype of the four individuals with ESRD, it lies outside the linkage peak generated from analysis of the whole family. Importantly, 1505 401 who had a renal biopsy in 1996 and continues to have proteinuria, does not carry the variant in *PCDH15*, while his phenotypically normal cousin, 1508 402, does. If 1508 401 does have the same condition as the rest of the family, then this is an informative meiosis, which would exclude *PCDH15* from the list of candidate genes.

DEAH (Asp-Glu-Ala-His) box polypeptide 32 *(DHX32)* is a novel gene with homology to the DHX family of RNA helicases. It is expressed in most cell types, and predominantly localised in the nucleus and mitochondria. The murine homologue, *Dhx32*, showed strong expression in the

mouse kidney, with significant protein expression in kidney tubules and little expression in the glomerulus [221]. The pattern of expression in various tissues was thought to suggest a cell-type dependent mechanism regulating Dhx32 expression. There is also evidence to suggest that *DHX32* expression correlates with human thymocyte maturation [222], and thus may play a role in lymphocyte differentiation [223].

Inositol polyphosphate 5-phosphatase F (*INPP5F*), also known as *phosphatidylinositide phosphatase SAC2*, is a 1,132 amino acid protein. It is ubiquitously expressed, but especially abundant in kidney, brain, heart, and skeletal muscle [224]. It exhibits 5-phosphatase activity specific for phosphatidylinositol 4,5-bisphosphate and phosphatidylinositol 3,4,5-triphosphate. However, unlike other 5-phosphatases, such as oculocerebrorenal syndrome of Lowe (*OCRL)* and inositol polyphosphate-5-phosphatase *(INPP5E)* that have well established associations with renal disease [225, 226], *INPP5F* does not contain the two conserved motifs that are essential for 5-phosphatase activity. The variant we had uncovered was a threonine to serine substitution at the C-terminal of the protein, well away from the SAC domain where the phosphatase activity is believed to reside.

Inositol and phosphatidylinositol phosphates are important for numerous cellular processes, including neuronal survival and signal transductions from growth factors, neurotransmitters and G protein coupled receptors. *INPP5F* also modulates the v-akt murine thymoma viral oncogene homolog 1 (Akt)/glycogen synthase kinase 3 beta (GSK-3β) pathway by decreasing Akt and GSK-3β phosphorylation [227, 228]. It appears to play a role in the developing mouse brain, as well as in the response to cardiac hypertrophy. The mouse with homozygous *inpp5f* knockout was not reported to show a renal phenotype; however the cardiac phenotype under study was mild, and required exogenous challenge to manifest itself [227].

4.3.08 Urothelial Cell Culture

As discussed earlier, the clinical phenotype in this family is of progressive renal impairment with minimal contribution from proteinuria and microscopic haematuria. This description would fit with a broadly tubular phenotype. Consequently I sought to look at the expression of these candidate genes in renal tissue. The most direct way of doing this would be to look at renal biopsy tissue from this family, however none was available to me, and moreover all the biopsies obtained to date had been fixed in formalin. I therefore sought to look at gene expression in cultured renal tubular cells obtained from a number of individuals in this family.

As part of normal physiology, approximately 2000 to 7000 cells from the human urinary tract detach and are excreted in the urine each day [161]. The majority of these cells come from the extensive network of renal tubules, whose total surface area is greater than that of the skin, rather than the downstream parts of the urinary tract, including ureters, bladder, and urethra. These cells, hereafter termed urothelial cells, are not damaged, but are fully functional and can be used for in vitro studies.

The efficiency with which urothelial cells can be cultured is quite variable, and depends on a number of factors, such as urinary pH and retention time, as well as on individual factors such as age and medication use [161, 229, 230]. Urothelial cells give rise to two main types of morphology, which may correspond to differing renal origins [229]. Urothelial cells also have limited lifespan in culture (1-6 passages), which makes it very hard to perform functional studies.

In order to validate this model, I first showed that these cells expressed epithelial markers that confirmed that they were indeed of urothelial origin. A range of urothelial markers have been used, mainly for immunohistochemical use or by stem cell biologists. I designed primers for three genes, E-cadherin (*CDH1),* keratin-7 (*KRT7*), and L1 cell adhesion molecule (*L1CAM*), and studied expression relative to a control population of primary dermal fibroblasts. *E-cadherin* is a classical marker of the epithelial lineage, first expressed at the 2-cell stage of embryogenesis. *KRT7* is a cytoskeletal protein found in simple glandular epithelium, and used as a histological marker of epithelial origin in urothelial cancer [229, 231]. *L1CAM* is a cell adhesion marker that is predominantly expressed in the late connecting tubule and collecting duct [232].

I cultured urothelial cells from a number of individuals in this pedigree (Figure 4.6). Of note, one individual (1508 304) had a functioning renal transplant at the time of collection, and by performing tetra primer PCR I was able to show that three separate urothelial cell colonies from this individual were all derived from the transplant donor and not from the individual in question, thus confirming the origin of these cells in the upper renal tract. Analysis of gene expression showed that all urothelial cell lines were derived from the renal collecting duct.

4.3.09 Expression of candidate genes in cultured cells

Detailed functional analysis has now become the most challenging and time-consuming stage of the investigation of novel genetic causes for human disease. In the absence of characteristic

findings on clinical or biopsy examination in this family, we had no *a priori* reason to choose between the three candidate genes identified by next generation sequencing. I therefore designed primers to look at the expression of all three genes in cultured urothelial cells (Figure 4.7).

There was no difference in the expression of any candidate genes in urothelial cells from one affected patient in this pedigree. Notably however, both *DHX32* and *INPP5F* are expressed at higher levels in urothelial cells than fibroblasts, compared to *PCDH15*. This would suggest that *DHX32* and *INPP5F* are important in this cell type, and consequently that alterations in function might lead to disease.

There are however a number of problems with this experiment. First, these cells showed a wide variety of gene expression, which may be consistent with the fact that they were a primary cell line, and not a homogenous tumour cell line. Second, although these cells express markers consistent with their urothelial origin, I was not able to identify their precise tubular origin. Notwithstanding the role that many other cell types, such a glomerular cells or fibroblasts, play within the kidney, it is therefore possible that I was studying the wrong cell population. Third, by sampling adult patients I may have been investigating the wrong developmental time point. Finally, and perhaps most importantly, it is very possible that a non-synonymous coding variant may have absolutely no effect on transcript expression. The next stage of analysis would therefore be to look at the effect of these coding variants on protein expression and function.

4.4 Conclusions and future work

A number of genes that are critical to kidney function in humans have been discovered in the last few years through the investigation of inherited renal disease [23-25]. I have studied a large Greek-Cypriot pedigree in order to discover another such gene. This family contains four individuals who developed ESRD at a young age and a large number of individuals with evidence of milder renal disease. Linkage analysis revealed a peak on chromosome 10 that did not include any genes known to be associated with renal disease. Given the size of the linked interval, I arranged for whole genome sequencing of two first cousins once removed. After appropriate filtering I was left with six novel heterozygous variants. Three of these variants occurred with significant frequency in the Turkish-Cypriot population and could therefore be excluded.

In order to study the effect of these variants *in vitro*, I cultured primary cells from individuals in this pedigree. I used urothelial cells derived from the renal collecting duct to look at gene expression, but was unable to find any effect. Urothelial cells are easily available; however they do not persist for long in primary culture. Dermal fibroblasts are a much more durable source of primary cells and would offer an alternative approach to studying the effect of these variants. Alternatively, urothelial cells could be immortalised, or transformed into induced stem cells, allowing long-term culture and the possibility of studying renal development *in vitro* [161].

Two candidate variants remain at this stage of the investigation, affecting *DHX32* and *INPP5F*. Given the lack of distinctive phenotype in the family it is difficult to know which may be responsible. *DHX32* is strongly expressed in renal tubular cells, and may play a role in cell differentiation through effects on translational machinery. *INPP5F* has been shown to modulate the Akt/GSK-3β pathway, which may be more amenable to study in cultured cells. However, neither variant is predicted to be especially pathogenic using current bioinformatic algorithms.

The difficulties faced by this study are illustrative of the limitations associated with recent advances in next generation sequencing technology. The main assumptions I have made are that the genetic mutation is shared by all four individuals with ESRD in this family, and that it is exonic. I have obtained a large quantity of high quality sequencing data, which has enabled me to show that renal disease co-segregates with just two novel variants in this family.

However, in the absence of a distinguishing phenotype or candidates with obvious mechanistic impact it becomes very difficult to demonstrate how these variants might be responsible for renal disease.

The identification of recessive conditions by next generation sequencing is significantly easier than the identification of dominant conditions because the genome of any given individual has around 50 times fewer genes with two, rather than one protein altering allele per gene [73]. Moreover, there remains a significant failure rate in the emerging studies that have used next generation sequencing to identify causative mutations in a cohort of patients with a given clinical phenotype [23, 233]. The recent identification of mutations in coding variable-number tandem repeats causing medullary cystic kidney disease type 1 are indicative of the challenges facing mutation discovery using massively parallel sequencing [234]. Consequently, the explanation for these failures of exome sequencing may ultimately help with identifying the mutation in this family. Fortunately the sequence data we obtained for this family includes all the intronic data that I have excluded from analysis so far. Furthermore, detailed clinical phenotyping of younger members in the family may help to identify candidate genetic pathways which we are now able to study *in vitro*.

Chapter 5 Mitochondrial kidney disease

5.1 Introduction

5.1.01 Mitochondrial disease

Linkage analysis is a robust tool for identifying the genetic basis of familial disease; however it assumes the manner of inheritance is understood. SNP-based linkage studies such as those discussed in chapter 4 are well suited to investigation of dominant or recessive autosomal conditions. However parametric linkage analysis requires that the correct mode of inheritance is specified, and analysis of sex-linked disease requires a different model. In addition to maternal inheritance due to X-linked disorders, a rare group of maternally inherited conditions are due to mutations in mitochondrial DNA (mtDNA).

Human mtDNA was first sequenced in 1981 [235], and the first pathogenic mutations were described within a decade. The mitochondrial genome is circular double-stranded DNA molecule of 16.6 kb, which encodes 13 proteins of the respiratory chain and 24 RNA genes [236]. More than 200 molecular defects have subsequently been described in patients with mitochondrial diseases [141, 237]. Point mutations can affect the various structural subunits of the respiratory chain, or compromise protein synthesis through the RNA genes. Alternatively large scale deletions of mtDNA remove one or more essential genes.

There are three major differences between mtDNA and nuclear genomic DNA. First, mammalian cells contain many copies of mtDNA, ranging from a few hundred to hundreds of thousands, depending on the cell type. Although the amount of mtDNA appears to be regulated in a tissue-specific manor, this can change over time through poorly understood regulatory mechanisms [238]. Second, pathogenic mutations can affect a variable proportion of the many mtDNA molecules, a situation known as heteroplasmy. The percentage of mutant mtDNA can vary between people, between tissues within an individual, and even between adjacent cells of the same individual [239]. Cells that contain a high percentage of mutant mtDNA express a biochemical defect, and this is often associated with a reduced amount of wild-type mtDNA [240]. Tissues and organs that contain many such affected cells may then give rise to the clinical features of disease. Third, mtDNA is exclusively inherited down the maternal line [143]. This means that men with mtDNA disease cannot pass on the disorder to their offspring. Women harbouring heteroplasmic mtDNA mutations pass on very different levels of mutant mtDNA to their offspring. This occurs because only a small group of maternal

mtDNA molecules are transmitted in oogenesis, resulting in rapid shifts in heteroplasmy levels within a generation due to statistical sampling effects [241, 242].

Mitochondrial disease typically affects the CNS and muscle tissue; possibly because both are post-mitotic and ectoderm derived. However there is good evidence for renal involvement in mitochondrial disease [243]. Thus there is a renal phenotype in some complex mitochondrial disorders, including Kearns-Sayre, Pearson, and Leigh syndromes [142]. There is an increased prevalence of deletions and SNPs in ESRD, however the extent to which these are contributory remains unknown [24]. Typically mitochondrial renal disease is associated with tubular and interstitial nephropathy and glomerulosclerosis.

5.1.02 Transcription of mitochondrial DNA

Transcription of human mitochondrial DNA is driven by a single subunit RNA polymerase and a set of transcription factors [244]. The two strands of mtDNA are differentiated by their nucleotide content with the guanine rich strand referred to as the heavy strand, and the cytosine rich strand referred to as the light strand [236]. In human cells each strand contains a promoter region for transcriptional initiation, the heavy strand promoter (HSP) and light strand promoter (LSP) (Figure 5.1) [245]. HSP transcription is initiated from two specific start sites, HSP1 and HSP2 [246, 247]. Both heavy strand promoters direct the synthesis of a poly-cistronic transcript which is subsequently cleaved by excision of the intervening tRNA transcripts to generate mature mRNA and rRNA molecules.

Intracellular levels of mature 12S and 16S rRNAs are in great excess to mRNAs [248]. This is now thought to be due to the action of the HSP1 site, which produces a transcript that terminates at the 3' end of the 16S rRNA gene [236]. *In vitro* measurement of HSP1 activity suggests that it is over 100 times stronger than the HSP2, however the HSP2 is responsible for the production of a full length heavy strand transcript [249]. Moreover the relative insensitivity of the *in vitro* system means that the exact role of each site *in vivo* remains unclear [244].

The replication of mtDNA is also carried out by unique enzymatic machinery encoded by the nuclear genome. Transcription from the LSP is not only necessary for gene expression, but it also produces the RNA primers required for initiation of mtDNA replication at the origin of H-strand DNA replication [236]. MtDNA replication occurs continuously throughout the cell cycle

[247]; however the switch between genomic length transcription and RNA primer formation and subsequent replication is not understood.

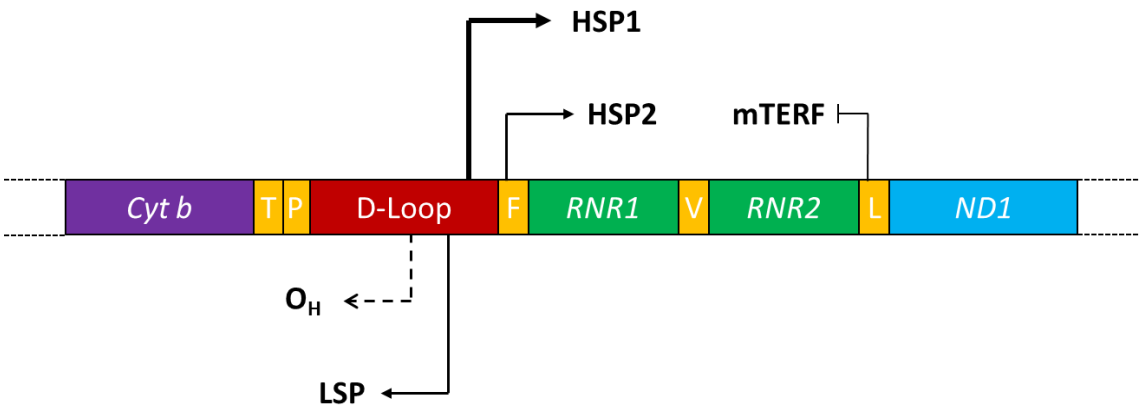


Figure 5.1 The mitochondrial promoter region

The mitochondrial D-loop region contains the three promoters that direct transcription of the mitochondrial DNA (mtDNA) heavy and light strands (HSP1, HSP2, and LSP), as well as the origin of heavy strand replication (O_H). Transcription from the HSP1 produces a short transcript encompassing the two ribosomal RNAs (shown in green), *RNR1* and *RNR2*, and two tRNA molecules (shown in orange) for phenylalanine and valine (F and V). This transcript is terminated by the action of the mitochondrial transcription termination factor (mTERF) which binds with high affinity to a site within the leucine tRNA (L). HSP2-initiated transcription produces the remaining mitochondrial mRNAs by processing nearly the full length of the mtDNA, including the genes for NADH-dehydrogenase subunit 1 (*ND1*) and cytochrome B (*Cyt b*).

Moreover, the molecular mechanism controlling mtDNA copy number in a given cell remains unknown. There may be an origin of replication for mtDNA maintenance under steady state conditions that is distinct from the origin required for recovery after mtDNA depletion and for accelerating mtDNA replication in response to physiological demands [250]. Copy number may also be influenced by the packaging of multiple mtDNA molecules into nucleoids [236, 251]. Two mtDNA maintenance proteins, mitochondrial transcription factor A (TFAM) and the mitochondrial helicase Twinkle are known to correlate linearly with mtDNA copy number [252, 253]. Lastly, there is a link between copy number and the energetic requirement of the cell, hence the differences in copy number observed between different tissues [238], and between the same cells exhibiting different levels of reactive oxygen species [254]. This last may help explain the persistence of DNA within the mitochondria when, according to the widely held endosymbiant theory [255], so much of the original proteobacterial genome has been transferred to the nuclear genome [236].

5.1.03 Cellular energy production

The best understood role of mitochondria is the production cellular energy and the regulation of cellular metabolism. Adenosine triphosphate (ATP) is generated through the aerobic respiration of pyruvate and reduced nicotinamide adenine dinucleotide phosphate (NADH) by over a hundred proteins organised into five respiratory chain complexes on the inner mitochondrial membrane. Each pyruvate molecule produced by glycolysis is actively transported into the mitochondrial matrix where it is combined with coenzyme A to form CO_2, acteyl-CoA and NADH. The acetyl-CoA is the primary substrate to enter the tricarboxylic acid (TCA) cycle. The TCA cycle oxidizes the acetyl-CoA to carbon dioxide, and, in the process, produces reduced cofactors that are a source of electrons for the electron transport chain. Reducing equivalents can also be imported from the cytoplasm to feed into the electron transport chain. This establishes a gradient of protons within the inter-membrane space that is used to generate ATP.

In addition to supplying cellular energy, mitochondria are also at the centre of cellular redox-dependent processes [256]. A small percentage of electrons within the inter-membrane space may prematurely reduce oxygen, forming reactive oxygen species (ROS). Mitochondria both generate ROS that drive redox-sensitive events, and respond to ROS-mediated changes in the cellular redox state. Most cellular ROS are partially reduced forms of molecular oxygen that result from electron leak during normal respiration in the mitochondrial electron transport chain. There is now a growing appreciation of the role of ROS in modulating cellular processes through redox-dependent signalling [256]. The mitochondrion therefore plays a central role in modulating metabolic, energetic, and ROS responses to promote growth and survival or cell death in response to a wide range of stimuli.

Efficient mitochondrial function is therefore dependent on the coordinated function of both the nuclear and mitochondrial genomes. The vast majority of respiratory chain components are encoded by the nucleus [257]. Moreover all the proteins involved in transcription and translation of the mitochondrial genome are also encoded by the nucleus. Mitochondrial disease can therefore be due to mutations of either mtDNA or nuclear DNA.

5.2 Results

I have studied an English pedigree containing 16 individuals with severe renal disease over three successive generations (Figure 5.2). Twelve individuals were recruited in 1996-7 for linkage analysis using limited microsatellite markers; however this approach was unable to determine a linked interval. The pedigree at this time included five individuals with ESRD, and four further individuals with renal impairment. One individual (322 412) had died aged 19 with a diagnosis of Batten disease, one of the neuronal ceroid lipofuscinoses (NCL). The clinical diagnosis was confirmed with a skin biopsy in 1993 which showed characteristic curvilinear fingerprint profiles on EM, and was thought to be unrelated to the renal condition in the rest of the family.

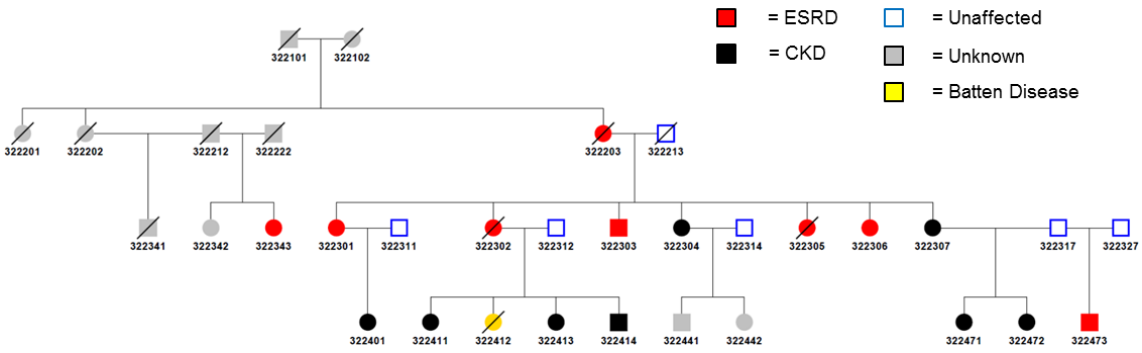


Figure 5.2 Pedigree of family 322

Individuals with evidence of renal impairment (affected individuals, filled shapes). Other individuals have either married in to the family (obligate unaffected, white shapes), or are phenotypically normal but may go on to develop renal impairment and thus were not included in the analysis (indeterminate, grey shapes).

In order to identify the mutation responsible for the renal disease in this pedigree I collected clinical details and DNA samples from an additional six individuals in this family. Of these individuals, two had reached ESRD and three others had abnormal renal function. The clinical phenotype in this family was notable for a lack of distinctive features. The newest member to reach ESRD had presented, aged 11, with renal impairment, and progressed to ESRD by age 16 with only mild proteinuria, no haematuria, normal renal tract imaging and autoantibody profile. In particular there was no evidence of renal tubular acidosis. He had declined a renal biopsy, but his mother (322 307) had one in 2010. Figure 5.3 shows the appearance of renal biopsies from two individuals in family 322 and Table 5.1 lists the main findings in all the renal biopsies performed on individuals from this family.

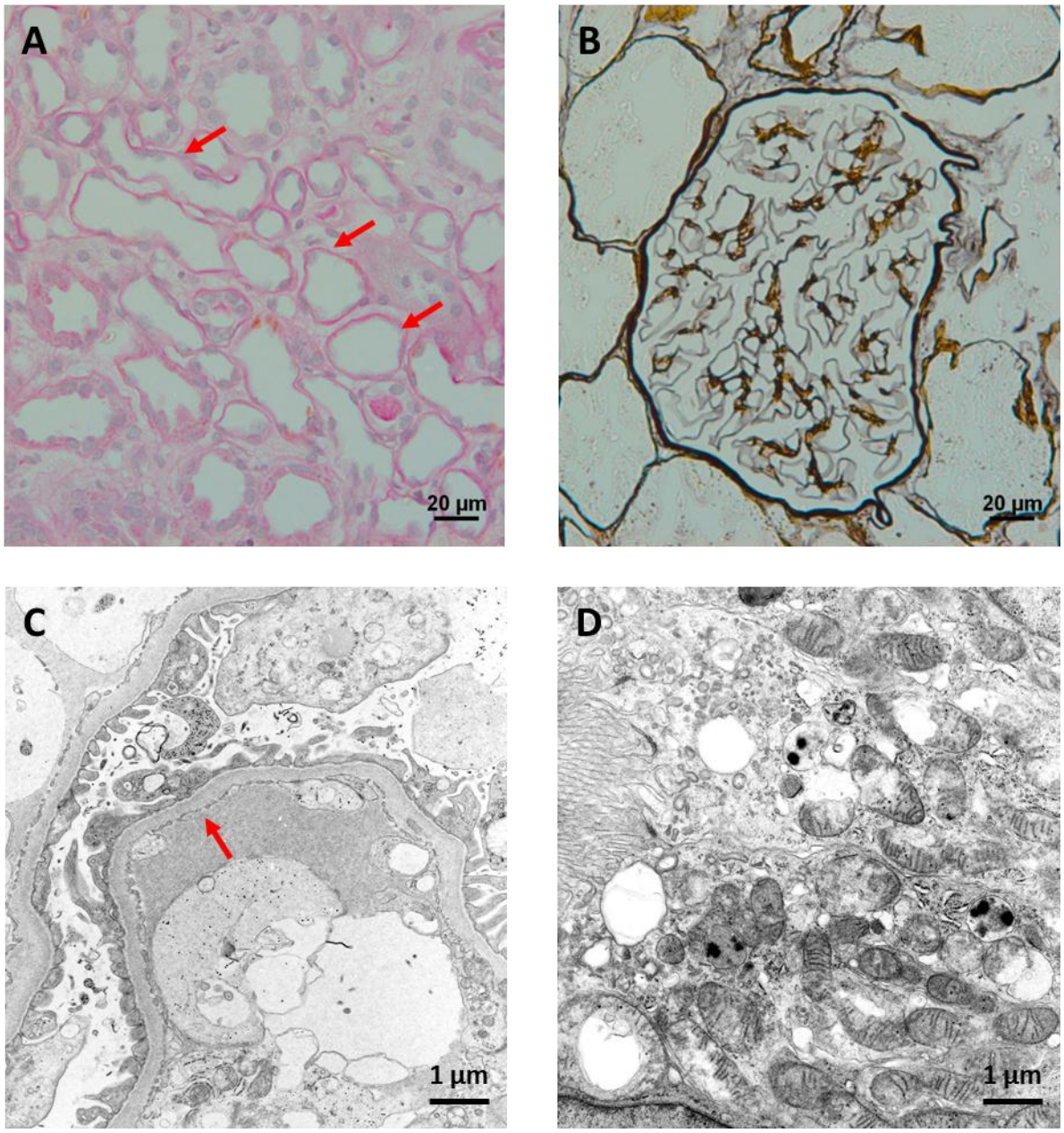


Figure 5.3 Renal biopsies from family 322

Light microscopic appearance of the renal biopsy from 322 303 showing an area of focal tubular atrophy that stains strongly periodic acid Schiff (PAS) positive (arrowed, A) and a glomerulus with normal appearance by silver stain (B). EM of the renal biopsy from 322 411, showing normal thickness and appearance of the GBM (arrowed, C) and normal appearance of tubular mitochondria (D).

ID	Date	Age	Location	Findings
303	1987	30	Oxford	Small and fragmentary with only 3 glomeruli No glomerular abnormality, IF not done, no deposition by immunoperoxidase but occasional flecks of C3 Focal tubulo-interstitial damage with tubular atrophy Tubular BM strongly PAS positive in atrophic areas EM: not performed
411	1994	22	Oxford	2/9 glomeruli with global sclerosis Immunofluorescence negative 2 more sclerosed glomeruli on frozen section Interstitial chronic inflammation and fibrosis with tubular damage EM: normal GBM, no immune deposits, normal mitochondria
304	1995	36	Leicester	2/12 with global sclerosis; otherwise normal appearance by light microscopy Two small areas of tubular atrophy No IF, but immunoperoxidase normal EM: single glomerulus with GBM of normal thickness but exhibiting odd irregular opacities in place of the lamina densa
301	2010	59	Oxford	Renal transplant biopsy (23y post-transplant) Widespread and focally severe acute tubular necrosis Little background chronic damage
307	2010	43	Leicester	6/16 glomeruli showing global sclerosis; otherwise normal appearance by light microscopy Focal tubular atrophy with fibrosis No IF, but immunoperoxidase normal Congo red stain for amyloid negative EM reprocessed from paraffin block but GBM thinned in areas (175 to 237 nm)

Table 5.1 Results of renal biopsies in family 322

Seventeen DNA samples were genotyped using the Illumina HumanCytoSNP-12 chip, giving nearly 300,000 SNP genotypes per sample. Preliminary analysis was performed using a subset of 9500 SNPs in order to reduce the analysis time. Genotypes were analysed using PedCheck [157], which detected the occurrence of nine Mendelian errors. These were discarded along with 22 genotypes which were thought to be 'unlikely' and may correspond to miscalled SNPs on the chip. The remaining genotypes were analysed using GENEHUNTER [154], using the EasyLINKAGE interface [153]. The genotypes were divided into sets of markers (80, 100, or 120) on each chromosome and parametric linkage analysis was performed in all individuals on each set, assuming a dominant model of inheritance.

A number of different definitions of affection status were used, such as only those with ESRD or those with CKD in one branch of the family, in case there were phenocopies within the pedigree. Using GENEHUNTER, analysis of the most basic pedigree comprising just those with ESRD and obligate carriers gave a LOD score of 2.06. Analysis of the extended pedigree including all those with elevated creatinine produced a LOD score of 2.64. Importantly, due to the size of the pedigree, GENEHUNTER was forced to ignore random individuals from analysis. Re-analysis of the linkage peaks on chromosomes 4, 10 and 15 using SimWalk gave a maximal LOD score of 1.86. Haplotype reconstruction demonstrated that no single chromosomal region was shared by all individuals with ESRD or with elevated creatinine, as shown in Figure 5.5.

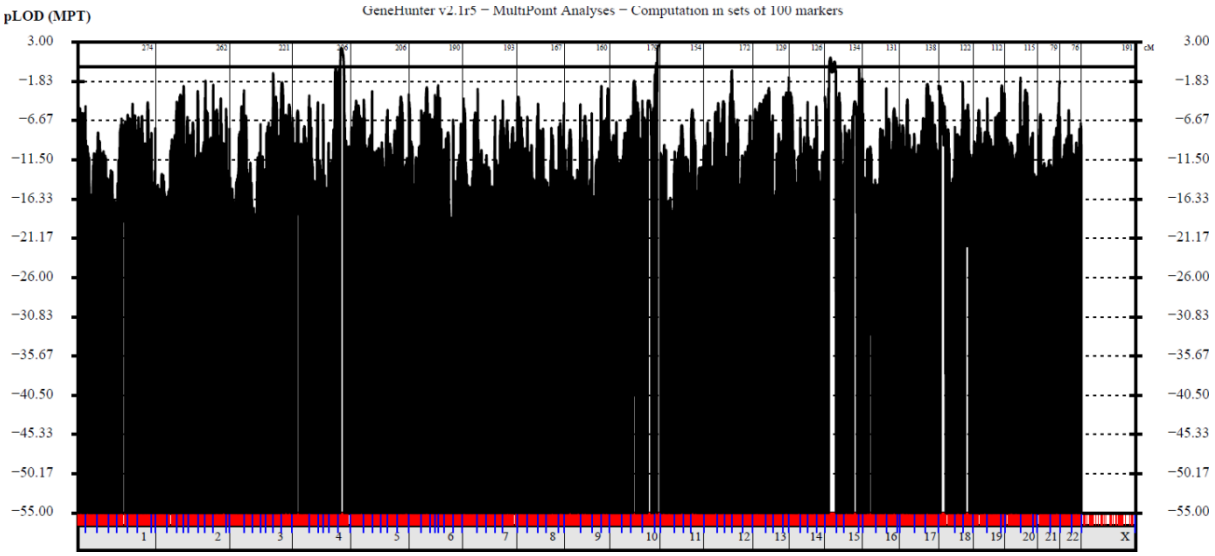


Figure 5.4 Linkage analysis of family 322

The maximal LOD score using Genehunter was 2.64 in association with a region on chromosome 10. Haplotype analysis revealed that some individuals were discarded by the programme, and repeat analysis using SimWalk including all individuals showed that this maximal LOD score was only 1.86.

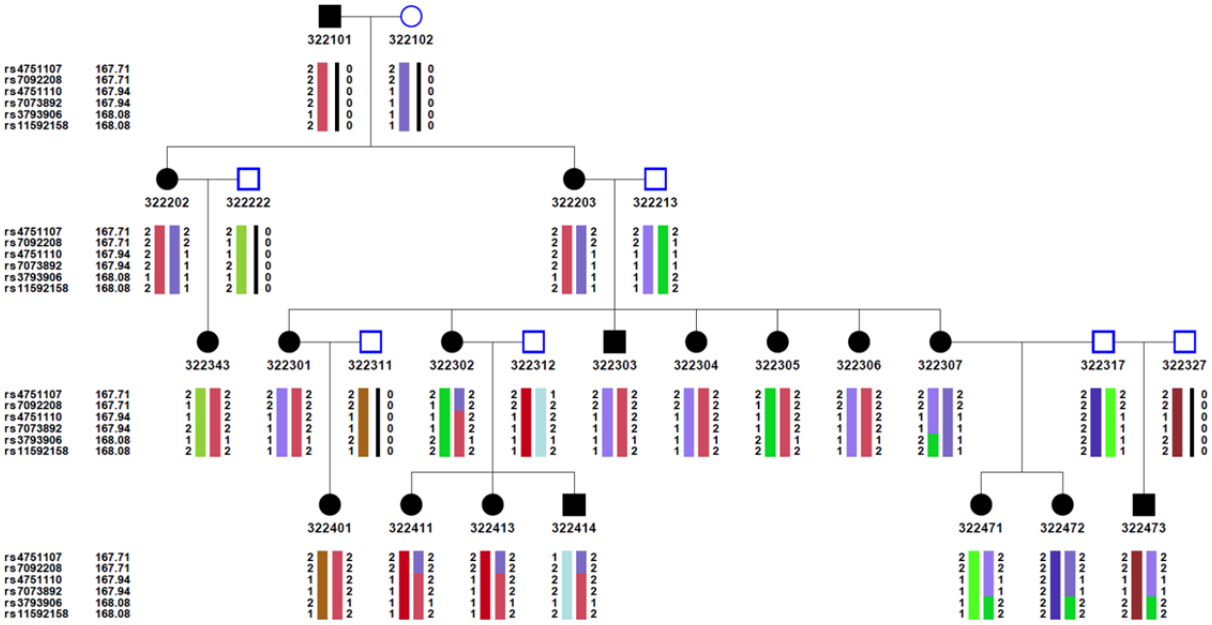


Figure 5.5 Haplotype of family 322 at chromosome 10 linkage peak

The maximal probability haplotypes generated by SimWalk for the linked region on chromosome 10 were visualised using Haplopainter to show the centromeric recombination.

The pedigree shown in Figure 5.2 is also consistent with maternal inheritance, as all affected individuals are related in the maternal line, and there is no male to female transmission. Parametric analysis assuming an X-linked model of inheritance did not show significant linkage. I was also unable to demonstrate a shared haplotype on the X-chromosome using conflicting homozygosity analysis.

I next proceeded to complete sequencing of the mitochondrial genome. Fresh urine samples were collected from four individuals in this family to obtain DNA for mtDNA sequencing. Sequencing was kindly performed by Dr Hudson at Newcastle University. Table 5.2 shows the non-synonymous variants found in all four affected individuals from this pedigree. The HSP1 variant was the only one not reported previously.

Variant	**Location**	**Amino acid**	**MAF**
m.547A>T	Heavy strand promoter 1 (HSP1)	NA	Unknown
m.8508A>G	*ATP8*	Asn>Ser	0.011
m.8860A>G	*ATP8*	Thr>Ala	0.022
m.10398A>G	*ND3*	Thr>Ala	0.459
m.13780A>G	*ND5*	Iso>Val	0.013
m.14178T>C	*ND6*	Iso>Val	0.020
m.14766C>T	*CYTB*	Thr>Iso	0.256
m.15326A>G	*CYTB*	Thr>Ala	0.008

Table 5.2 Non-synonymous mitochondrial sequence variants found in family 322

The m.547A>T polymorphism introduces an Mbo1 restriction site into the sequence that can then be detected by RFLP analysis, allowing precise quantification of the mutational load using labelled primers. All individuals in the maternal line of this family share the HSP1 polymorphism at homoplastic levels. This variant is also amenable to analysis by tetra-primer PCR, as shown in Figure 5.6. In order to determine the sensitivity and specificity of this assay for mitochondrial DNA I prepared serial dilutions of patient and control DNA. The abundance of mitochondrial DNA means that this assay could be performed with 1000-fold dilutions of DNA, far below the detectable limit of spectrophotometric detection.

Tetra-primer PCR was performed on DNA extracted from blood leukocytes, urinary epithelial cells and saliva in three individuals with ESRD from family 322. There was no change in the levels of mutant mtDNA in these different tissues. By comparing DNA collected from leukocytes in 1996 and 2011 it was also apparent that there was also no change in the levels of mutant mtDNA over time.

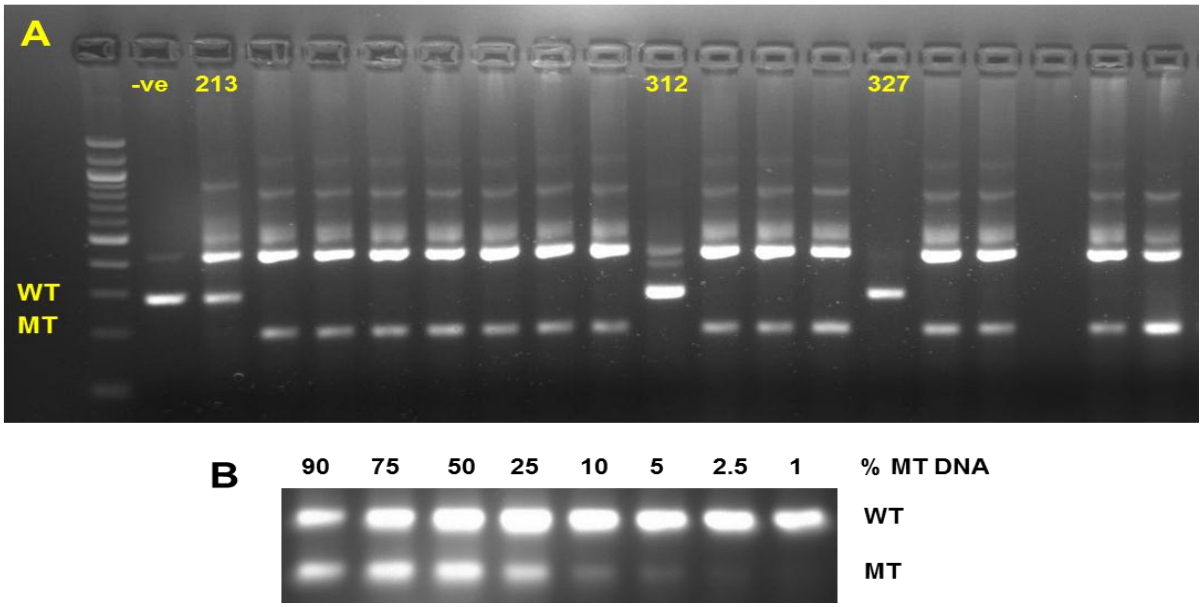


Figure 5.6 Tetra-primer PCR analysis of HSP1 polymorphism in family 322

(A) Occurrence of the m.547A>T polymorphism at homoplastic levels in all individuals descended in the maternal line of family 322. Spouses married in to the family (213, 312, and 327) all show the wild-type PCR product. (B) Sensitivity analysis for the tetra-primer PCR reaction in detecting mitochondrial heteroplasmy. The mutant product is detectable at <5% of starting DNA.

In order to investigate whether the m.547A>T or other HSP1 polymorphisms might be associated with renal disease in any other individuals I undertook sequencing of the entire HSP1 region in 46 patients with undiagnosed familial renal disease. Using the D4 primer pair I amplified nucleotides 323-752 of the reference sequence, which includes the end of the D-loop region. Table 5.3 shows the list of variants identified in this region in this cohort. No other individuals were found to have the m.547A>T polymorphism described in this family, or any other variants in the HSP1 region. Many other variants were observed, however, including those associated with mitochondrial haplogroups such as m.477T>C or m.750A>G (http://www.mtdb.igp.uu.se/). Rare variants were also observed, such as m.368A>G or m.523+CA, but the significance of these is not known.

Position	Region	Change	Individual	MAF	Phylotree
368	LSP	A>G	20	0.001	not present
455	D-loop	+T	9, 28	NA	Present on I1 (subclade of N)
462		C>T	18, 42	0.033	J1
477		T>C	32	0.009	H1
482		T>C	42	0.007	J1C
489		T>C	18, 42, 44	0.362	J
498		delC	39	NA	k1c
513		G>A	8	0.013	homoplastic
516		C>T	33	0.001	U5b1c2
523	TFAM	+CA	10, 25	NA	
533		A>G	8	0.001	H10/U5b1f - Homoplastic
573	D-loop	+C x7	28	NA	Common tract/repeat
574		delA	28	0.000	
575		C>T	13	NA	H/U - Homoplastic
709	RNR1	G>A	3, 11, 16, 20, 23, 30, 35, 36	0.164	homoplastic
750		A>G	All samples	0.992	All modern humans
751		A>G	13	NA	H2a1e1

Table 5.3 D-loop variants found in 46 patients with unidentified familial renal disease

In order to study the functional significance of this polymorphism, dermal fibroblasts were collected from four affected individuals and three of their healthy partners. Early passage fibroblasts were used for all experiments and genotype was confirmed by PCR (Figure 5.7).

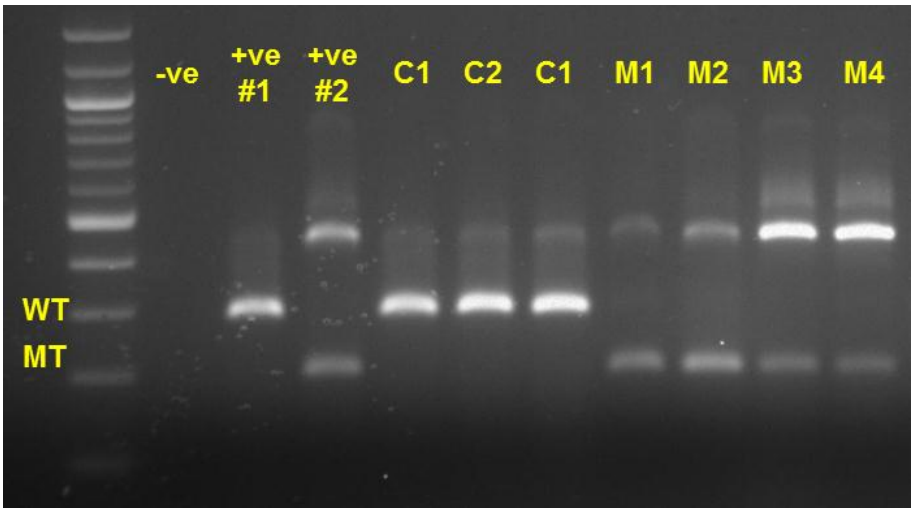


Figure 5.7 Genotype of primary fibroblasts

Tetra-primer PCR for the m.547A>T polymorphism shows that DNA extracted from the three control lines (C1-3) express the wild-type product (WT) at an equivalent level to DNA extracted from control leukocytes (+ve #1). The four patient-derived lines (M1-4) express the mutant product (MT) at homoplastic levels, equivalent to the urinary DNA used in the original sequencing experiment (+ve #2).

The function of the m.547A>T polymorphism was assessed in a number of ways, including analysis of gene expression and DNA replication. In order to distinguish between the relative importances of mitochondrial transcriptional promoters, I looked at gene expression relative to both nuclear and mitochondrial genes (Figure 5.8). There was a significant reduction in the expression of *RNR1* (under the control of the HSP1) when expressed relative to *MT-CO1* (under the control of the HSP2) (Figure 5.8). This effect was not seen with expression relative to a nuclear gene, or with the expression of *ND6* (under the control of the LSP) relative to *MT-CO1*.

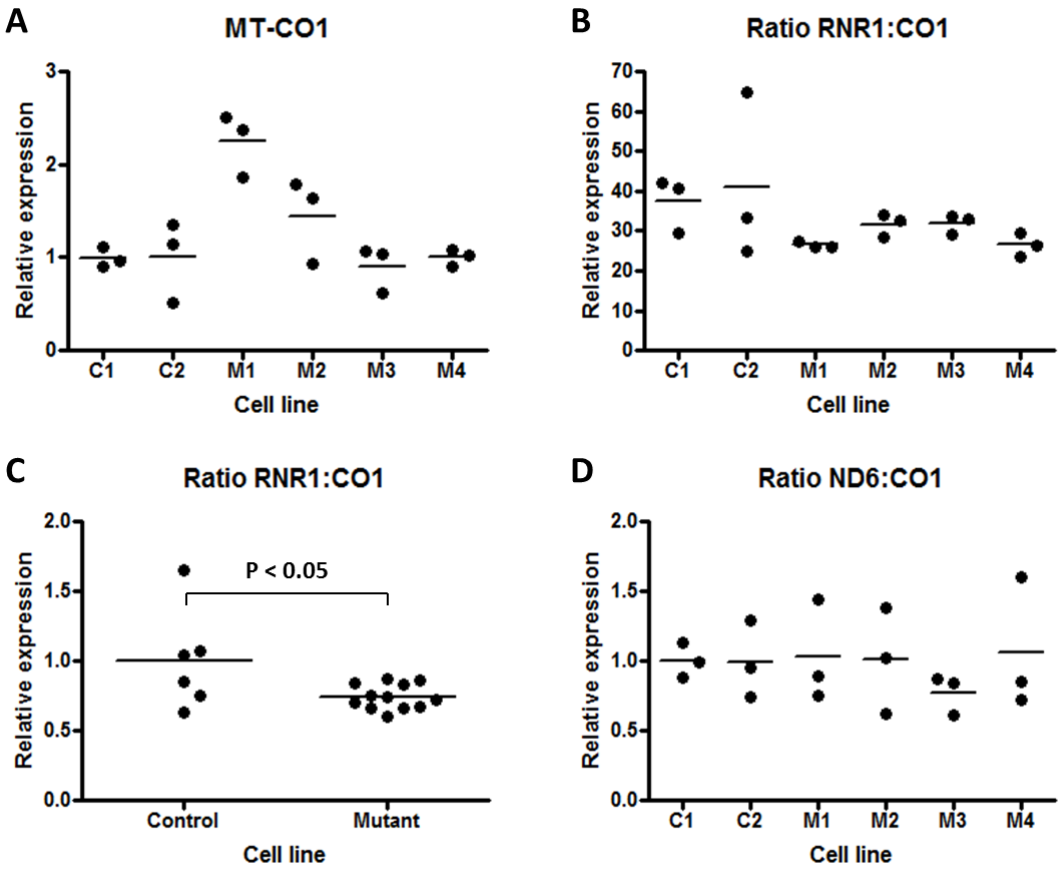


Figure 5.8 Analysis of gene expression in primary fibroblasts

Expression of genes under the three mitochondrial transcriptional promoters (HSP1, HSP2, and LSP) was assessed by real-time PCR. There was no effect of genotype on expression of MT-CO1 (A), however the expression of RNR1 (under control of HSP1) was reduced when expressed relative to MT-CO1 (under control of HSP2, B). This effect was significant when individual cell lines were pooled by genotype (C). This effect was not seen with ND6 (under the control of the LSP, D).

I also looked at expression of mitochondrial proteins (Figure 5.9). There was no effect of the HSP1 polymorphism on the expression *MT-CO1*, when compared the nuclear encoded mitochondrial proteins gene, VDAC1 and Cyt C. It is possible that reduced expression of mitochondrially encoded rRNAs would have an effect on the translation of *MT-CO1* mRNA. However, given that I had previously shown that levels of *MT-CO1* mRNA were not affected by the HSP1 polymorphism, it is not surprising that there was no effect on protein levels.

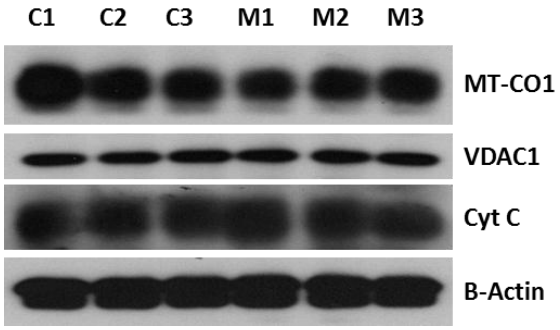


Figure 5.9 Analysis of protein expression in primary fibroblasts

Protein expression of the mitochondrially encoded MT-CO1 was compared to nuclear encoded mitochondrial genes, VDAC1 and Cyt C.

MtDNA copy number appeared increased in patient-derived fibroblasts (Figure 5.10). This difference was significant if control and mutant experimental groups were pooled together (Figure 5.10). However, without knowing the mitochondrial haplogroup of the control fibroblasts such comparison may not be valid [258].

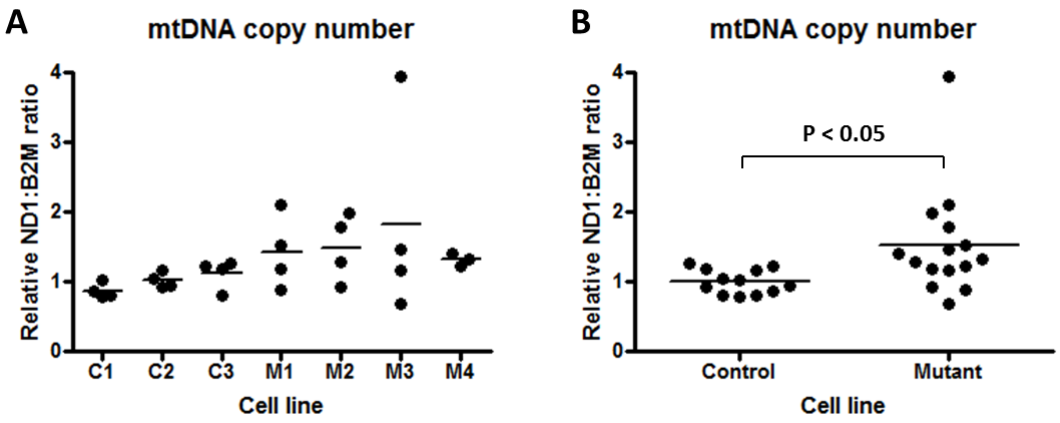


Figure 5.10 Analysis of mitochondrial DNA (mtDNA) copy number in primary fibroblasts

Mitochondrial DNA copy number was assessed in each cell line (A) and pooled (B) by measuring the relative expression of ND1 and a single-copy nuclear gene, B2M, using real-time PCR of total cellular DNA [259].

In addition to these measures of transcription and translation, patient-derived fibroblasts were used to study mitochondrial function *in vitro.* Control cells treated with ethidium bromide to deplete their mitochondria were included as a negative control. Figure 5.11 shows a representative trace of the oxygen consumption rate (OCR) measured with the Seahorse bioanalyser. Measurements of oxygen consumption and proton release in patient-derived fibroblasts revealed a marked effect of the HSP1 polymorphism (Figure 5.11, Figure 5.12).

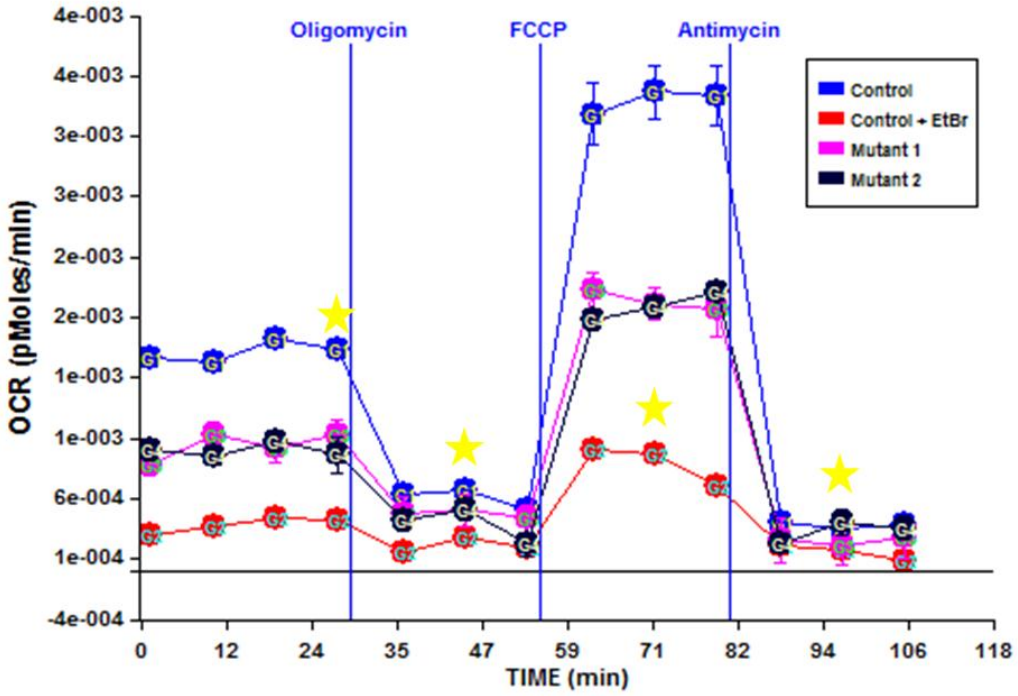


Figure 5.11 Oxidative phosphorylation in patient-derived fibroblasts

Oxygen consumption rate (OCR) was measured every 8 minutes, each data point shows the mean and standard error from 3 - 5 replicate wells, normalised for cell number. Inhibitors of the electron transport chain were added at the times indicated, and the measurements used for subsequent analysis are indicated by a yellow star.

There was a significant reduction in baseline OCR with a compensatory increase in the extracellular acidification rate (ECAR) (Figure 5.12). This was associated in mutant cells with a reduction in maximal respiratory rate (after carbonyl cyanide p-[trifluoromethoxy]-phenyl-hydrazone (FCCP) treatment), respiratory control rate (FCCP – oligomycin), ATP production (baseline – oligomycin), spare respiratory capacity (FCCP – baseline), and maximal respiration (FCCP – Antimycin A). However all of these differences disappeared if the results were first normalised to basal OCR rates. The only measure not affected in this way was the proton leak (oligomycin – Antimycin A), which was equal under baseline conditions, but significantly elevated by normalising to basal OCR (Figure 5.12).

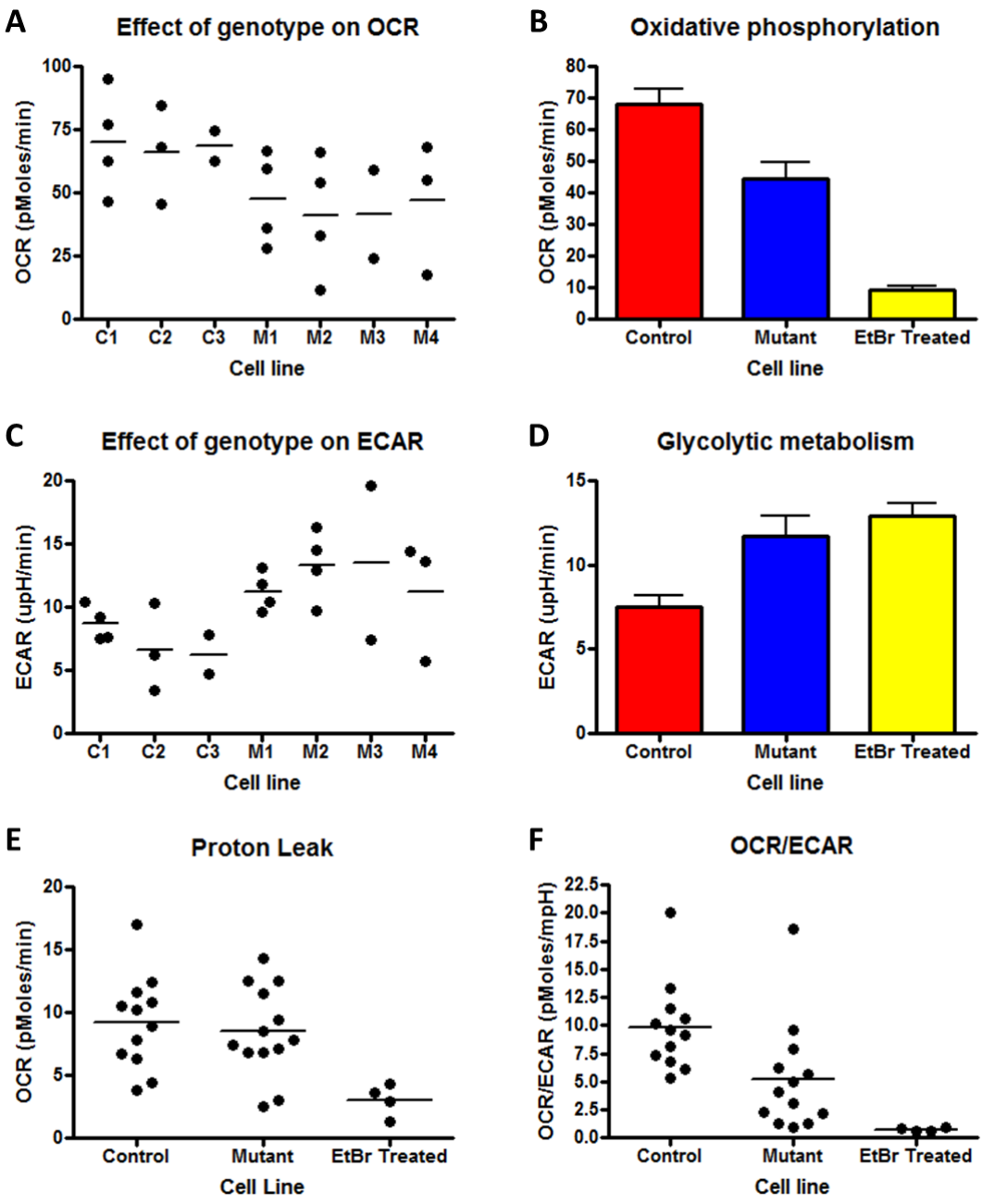


Figure 5.12 Analysis of mitochondrial metabolism in patient derived fibroblasts

Baseline oxygen consumption rate (OCR) measured for all individual cell lines (A) and pooled (B, $p < 0.0001$). Baseline extracellular acidification rate (ECAR) measured for all individual cell lines (A) and pooled (B, $p < 0.0005$). Measurement of proton leak (E), calculated as OCR after treatment with Oligomycin minus OCR after treatment with Antimycin (p = NS, but $p < 0.05$ if corrected for baseline OCR); Ratio of OCR to ECAR (F) showing mutant cells have intermediate phenotype ($p < 0.005$).

5.3 Interpretation

5.3.01 Failure of linkage analysis

As discussed in chapter 4, the presence of 17 affected individuals in this pedigree would give a maximal LOD score of 17 x 0.3 = 5.1. Consequently, if the kidney disease was due to a single genetic variant in this pedigree, linkage analysis should produce a significant LOD score. That this was not observed could be due to one of a number of problems with this type of analysis. First, renal disease could be due to the interaction of a number of shared genetic factors interacting with a common environment. However the chance of this occurring in an outbred white population is unlikely. Second, there might be single genetic factor, but the signal from linkage analysis be obscured by the presence of phenocopies in the pedigree. Thirdly, the assumption of autosomal dominant inheritance in this family may be erroneous.

It is difficult to exclude the presence of phenocopies, because the phenotype in this family lacks distinguishing features. As shown in Figure 5.5, if 307 and descendants were all phenocopies, this would explain the loss of linkage to this region of chromosome 10. In addition to linkage analysis I also assessed conflicting homozygosity. Using combinatorial analysis I was able to consider different permutations of individuals within the pedigree to see whether any groups of individuals shared a haplotype that was IBD. This analysis did not show up any regions that were IBD beyond that expected by chance alone. Consequently, it is unlikely that a small number of phenocopies were obscuring the signal from linkage analysis.

We therefore concluded that we were incorrect in our assumption of the mode of inheritance. The pedigree is also consistent with a maternally inherited disorder, but CH analysis also ruled out a shared haplotype on the X chromosome. MtDNA was not included on the first SNP chips used for identifying nuclear disease. Moreover the high rate of mitochondrial mutation means that it is not possible to analyse the mitochondrial genome in terms of linkage disequilibrium. I therefore proceeded to sequence the mitochondrial DNA in this family.

5.3.02 Mitochondrial sequencing and heteroplasmy

Due its small size it has been possible to sequence the mitochondrial genome for over 30 years [235]. However, the presence of heteroplasmy in most mitochondrial disorders renders the sensitivity of this test quite variable. Levels of a pathogenic mutation can vary both between individuals and between tissues [260]. While DNA is commonly obtained from peripheral

leukocytes for the assessment of nuclear DNA, it shows high levels of heteroplasmy for mtDNA. Leukocyte DNA may also show progressive loss of mutant mtDNA with time [261, 262].

Muscle tissue therefore remains the gold standard for the diagnosis of mitochondrial disorders. However muscle biopsies are invasive and there is evidence that suitable DNA can be extracted from other tissues, such as cultured fibroblasts and hair shafts [260, 263]. Importantly, DNA extracted from urothelial cells shows close correlation with muscle for the detection of common pathogenic polymorphisms. I therefore collected fresh urine samples from four affected individuals and one control in this family to obtain DNA for mtDNA sequencing.

Once we had details of the mtDNA sequence I designed a tetra-primer ARMS PCR reaction to study the HSP1 polymorphism. This PCR reaction is dependent on competition between two primers with near equal affinity for a DNA polymorphism [152]. With nuclear DNA the reaction can give one of three results: heterozygous or homozygous for either allele. However, given the nature of mitochondrial heteroplasmy, allele frequency exists on a continuum from 0-100%. The original Sanger sequencing and RFLP analysis that was performed on DNA from this family can be used to give an idea of heteroplasmy by looking at the relative intensity of dye incorporation, and informed us that this polymorphism was at homoplastic levels.

I examined the sensitivity of this PCR assay using a range of mixture of wild-type and mutant mtDNA. Figure 5.6 shows that it was sensitive to a 40-fold change in the degree of 'heteroplasmy' for this polymorphism. It also confirms the observation from Sanger sequencing that this polymorphism was at homoplastic levels in this family.

5.3.03 HSP1 sequencing and mitochondrial polymorphisms

One important route to determining the pathogenicity of a given polymorphism is to determine its prevalence in the general population and in a population with the same phenotype. One of the major issues with the comparison of mitochondrial sequences is the high degree of variation within the human population. There is less publically available data on mitochondrial than nuclear sequence variants, however none of the 5680 NCBI sequences had the m.547A>T polymorphism and three (including one European) had an m.547A>G polymorphism.

In order to investigate whether the m.547A>T polymorphism or any other variants of the HSP1 region might be associated with renal disease in any other individuals I undertook sequencing of the entire HSP1 region in 46 patients with undiagnosed familial renal disease. These 46 individuals came from pedigrees that included all modes of inheritance, where known, since families with apparently maternal inheritance are so scarce that no single centre is likely to have such a collection. The rationale for sequencing the entire HSP1 region was that if the m.547A>T polymorphism did indeed mediate its effect through mtDNA transcription then we might have a better chance of picking up other variants affecting promoter binding than just looking at a single SNP.

The mitochondrial haplogroup describes variations from the reference sequence that tend to be inherited together. Other sequence variants reported in this family place it within haplotype N, occurring in 1% of the Caucasian population in the UK, and representative of early European, Middle Eastern, or Finnish ancestry (http://www.mitomap.org/bin/view.pl/MITOMAP/HaplogroupMarkers).

5.3.04 Fibroblast studies

Dermal fibroblasts are widely used in studies of human genetic disease. They are easily cultured from a forearm skin biopsy, and are able to survive in culture for a number of passages. Skin biopsies typically contain a mixed population of cells, including keratinocytes and micro-vascular endothelial cells; however fibroblasts become the dominant cell type within a few passages in routine conditions.

Adaptation to growth in culture medium is an important limitation with the study of cultured fibroblasts. Fibroblasts become highly glycolytic soon after removal from the body when grown in glucose rich medium, such as that used in this study (4.5 g/L) [264]. High glucose itself is a mitochondrial stressor [265]. In addition to adaptations to culture media, some substrates are critical to the normal growth of cells harbouring certain mitochondrial mutations, namely pyruvate and pyrimidines [266, 267].

Early studies of osteosarcoma cells that lacked mtDNA showed that they were auxotrophic for pyruvate [266]. The leading explanation for this is that excess pyruvate is essential to prevent an excessive shift in the ratio of NAD to NADH in the absence of oxidative phosphorylation [268]. Cells with mitochondrial mutations are also auxotrophic for pyrimidines, which can be

overcome by supplementing cells with uridine [267, 269]. This is due to the coupling of a critical enzyme in the pyrimidine biosynthetic pathway, dihydroorotate dehydrogenase, located on the inner mitochondrial membrane to the activity of the electron transport chain. Consequently all studies were performed in growth medium supplemented with both pyruvate and uridine.

Ethidium bromide is a cationic mutagen capable of intercalating into DNA during replication. Because of its positive charge, it concentrates within negatively charged mitochondrial matrixes. This allows determination of cell culture concentrations (typically 50 ng/ml) that block mtDNA replication without disrupting nuclear DNA replication [266]. Prolonged exposure to ethidium bromide causes selective depletion of mtDNA in human fibroblasts [259, 268]. MtDNA levels can be repopulated if the duration of exposure is short (1-2 weeks), but may become irreversible, with attendant loss of expression of respiratory chain enzymes, such as Complex IV [266, 267]. I used control fibroblasts treated with ethidium bromide for 1 week as a negative control in the experiments in Figure 5.11.

There are two major limitations with using fibroblasts to study mitochondrial disease. First, they are quite distinct from the tissue which causes the human disease. Affected members of the pedigree under study do not have any skin phenotype that might be associated with mitochondrial disease. Instead this disease appears to be due to pathology of the renal tubule, which is derived from epithelial cells of the mesoderm. Second, fibroblasts continue to divide in culture, whereas a number of mitochondrial diseases are only manifest in terminally differentiated tissue, such as neurons or myotubes. These two limitations may help to explain the absence of significant findings seen in experiments using primary fibroblasts.

5.3.05 Mitochondrial gene expression

Mitochondrial transcription results in the production of a single poly-cistronic transcript that is subsequently cleaved to yield individual RNA molecules. The HSP1 site is significantly more active than the HSP2, and results in the synthesis of a transcript encompassing the first four RNAs, namely the two ribosomal RNAs, RNR1 and RNR2, and two tRNA molecules (F and V) (Figure 5.1). I therefore examined the expression of transcripts under the control of HSP1 (*RNR1*) and HSP2 (*MT-CO1*) relative to each other, as well as relative to a nuclear housekeeping gene (*B2M*). It is not possible to assess the level of the tRNAs by real time PCR

due to their complex secondary structure. However given the poly-cistronic transcript, these ought to be expressed at a comparable level to *RNR1* and *RNR2*.

An additional difficulty with the mitochondrial RNA transcript is that it is not possible to design RNA specific primers, due to the lack of introns. In order to overcome this, I included an additional DNase digestion step to the RNA isolation protocol. Mitochondrial mRNAs undergo significant post-transcriptional regulation [270]. However, stable polyadenylation of mt-mRNA 3'-termini is a common feature, and would have been recognised by the oligo (d-T) primers included in the retrotranscription kit.

Given the high abundance of mitochondrial transcripts assessed by PCR, the effect of this promoter could be very small. This may explain why there was no consistent effect of genotype on the expression of either RNR1 or other heavy strand transcripts. The reduction in the ratio of *RNR1* to *MT-CO1* expression is intriguing, however. Decreased activity at the HSP1 would be predicted to reduce transcription of the shorter transcript, while leaving full length transcription under the control of the HSP2 unaffected. Whilst this effect was small, and only significant when the cell lines were pooled by genotype, it is possible that this may have a significant effect *in vivo*.

5.3.06 Mitochondrial protein expression

The vast majority of mitochondrial proteins are encoded by the nucleus. I therefore looked at the expression of both nuclear and mitochondrially encoded proteins. I studied mitochondrially encoded cytochrome c oxidase 1 (*MT-CO1*) and the nuclear encoded voltage-dependent anion-selective channel protein 1 *(VDAC1*) and cytochrome C (*CYCS*). MT-CO1 is one of three mitochondrially encoded subunits of the complex IV. VDAC1 forms a channel through the mitochondrial outer membrane that allows diffusion of small hydrophilic molecules. There was no effect of the m.547A>T polymorphism on mitochondrial protein expression.

Comparison of mitochondria across different tissues shows significant correlation between protein levels and RNA levels [271]. This suggests that, despite the importance of post-transcriptional regulation, mitochondrial protein mass is controlled largely at the level of transcription [272]. Consequently the failure to detect significant changes in protein expression is not surprising given the small difference at the mt-mRNA level. It remains

possible that a subtle difference exists that might be shown by serial protein dilution or careful quantitation of band intensity.

5.3.07 Mitochondrial DNA copy number

The regulation of mitochondrial copy number is poorly understood, despite the importance of this process in human physiology and disease [141, 273]. MtDNA replication occurs continuously, regardless of growth state, in both proliferating and post-mitotic cells [247]. Loss of mtDNA in neurones of the substantia nigra has been shown to be a common cause of neurodegenerative conditions, and defects of mtDNA maintenance may have a role in the mechanism of aging [273]. Of the proteins required for mitochondrial replication, only TFAM and Twinkle have been shown to increase copy number in vivo, when expressed at physiological levels [252, 253]. Increasing the amount of mtDNA has been suggested as a therapeutic approach in diseases with mitochondrial dysfunction.

There is also a link between mtDNA copy number and oxidative stress that has been observed in human cells [274, 275]. Using a murine cybrid model, defective mitochondria were associated with increased mitochondrial biogenesis and increased generation of reactive oxygen species (ROS) [254]. In this paper the increased mtDNA copy number was abolished by treatment with antioxidants such as N-acetyl-L-cysteine (NAC) and Tiron.

This implies that the increased mtDNA copy number observed in patient-derived fibroblasts is compensation for a reduction in some other aspect of mitochondrial function. As shown previously, the net total level of gene and protein expression is unaffected; however the expression per copy of mtDNA is reduced. This might result in an overall change in bioenergetics function or free radical production, which would be even greater in a tissue with higher mtDNA copy number. It is not possible to control the copy number using cultured cells, however such experiments may be possible in vitro using a recombinant *in vitro* transcription system [244]. An alternative experiment would be use trans-mitochondrial cybrids to control for nuclear background.

5.3.08 Oxygen consumption rate

In parallel with these measures of transcription and translation, I employed patient-derived fibroblasts to study mitochondrial function *in vitro* using the Seahorse XF bioanalyser [276,

277]. This instrument provides real time measurement of the extracellular flux of both oxygen and protons. Using a probe to create transient 'microchambers' about the cells in culture, it uses fluorescent sensors to measure oxygen consumption rate (OCR) and extracellular acidification rate (ECAR). Up to 20 assay chambers are usable in one assay allowing internal replication and direct comparison of multiple cell lines or conditions.

Fibroblasts from patients with the m.547A>T polymorphism have a significantly reduced oxygen consumption rate compared to controls. OCR was reduced under basal conditions in all four patient-derived cell lines, when compared with three control lines that had been collected at the same time. I used oligomycin to inhibit ATP synthase, giving a measure of ATP production, FCCP, an un-coupler that enables measurement of maximal respiration, and Antimycin, which inhibits complex III, giving a measure of underlying proton leak and non-mitochondrial respiration. These measures were all significantly lower in patient-derived cells, however this difference was abolished if rates were normalised for baseline OCR.

Measurement of proton efflux (ECAR) shows that not only are patient cells less oxidative, but they are also more glycolytic. This would enable the cells to maintain the same energy balance as control cells, and explains why these cells appeared to behave identically in culture to control fibroblasts.

One limitation of these results is that prolonged growth in culture is associated with a progressive decline in respiratory capacity as measured using the Seahorse XF bioanalyser. This may represent a switch to glycolysis, as discussed previously, or alternatively it may reflect cellular senescence or some other mechanism entirely. Using a whole cell fluorescent measure of ATP, as much as 85-90% of ATP generation in these cells is unaffected by treatment with oligomycin (data not shown). Irrespective of the mechanism, a skin biopsy required culture for 3-4 weeks before there were sufficient cells to analyse.

Although these cells exhibited reduced OCR under standard culture conditions in high glucose media, it is hard to explain mechanistically how this has come about. I utilised specific inhibitors for complex III, IV, and V, which did not reveal any differences in in function once baseline had been corrected for. It is also possible to use specific substrates for complex I, II, and IV, which should correct loss of function at these earlier stages of the electron transport chain [254]. Previous studies of mitochondrial cybrids containing a tandem duplication of the HSP region showed no effect on oxidative phosphorylation [278], however it appears that

larger deletions including the HSP region show abnormal mitochondrial proliferation and COX staining *in vivo* [279]. It remains unclear, therefore, why this promoter variant has such a significant effect on mitochondrial function.

5.3.09 Genetic factors affecting mitochondrial metabolism

There are two additional considerations that may be relevant to the differences in oxidative metabolism shown by patient fibroblasts. First, the diagnosis of Battens disease may not be entirely unrelated to the disease seen in the rest of the family. There is evidence of mitochondrial involvement in this condition, notably with evidence of morphological and functional alterations in both human and animal NCL fibroblast cultures [280, 281]. In the English setter model of disease, impaired mitochondrial function was noted in heterozygous animals [282]. While I have shown that no single autosomal region can account for renal disease in this family, it is possible that a heterozygous NCL mutation might be contributing to the mitochondrial phenotype observed in the fibroblasts from the four individuals studied here. Conversely it is possible that the label of Batten disease was in fact a *forme fruste* of the mitochondrial disease evident in the rest of the family.

Second, there is some work now suggesting that mitochondrial haplogroup can also influence oxidative phosphorylation. Interest in the field was triggered by epidemiological studies showing differences in the risk of developing mitochondrial diseases, such as Leber's hereditary optic neuropathy (LHON), in different haplogroups. Subsequent research was conducted in cybrid models, enabling the researcher to reduce the influence of nuclear DNA seen in primary cell cultures. No detectible differences in respiratory function were observed when mtDNAs belonging to European haplogroups X, H, T and J were used in a study by Carelli *et al* 2002 [283]. However, more recently haplogroup J cybrids were shown to display lower oxygen consumption, mitochondrial inner membrane potential and total ATP levels than haplogroup H [258]. Moreover there is good evidence to suggest that variations in mtDNA influence the response to various drugs [284]. Given that the haplogroup of this family is rare in the UK, it is conceivable that the observed difference in oxidative metabolism is due to differences in haplogroup between patients and controls.

5.4 Conclusions and future work

This chapter describes the investigation of another large pedigree with multiple members affected by renal disease across three generations. Linkage analysis was not consistent with inheritance of a mutation in nuclear DNA, and the pedigree is consistent with a maternally inherited disorder. Subsequent sequencing of the mitochondrial genome demonstrated a novel polymorphism of the HSP region that was thought to be potentially pathogenic.

I have cultured primary dermal fibroblast derived from individuals in this pedigree with renal disease. Fibroblasts demonstrated two abnormalities; first, an increase in the mtDNA copy number, and second, a reduction in oxidative respiration with an associated increase in glycolysis. This occurred in the absence of changes in mitochondrial gene or protein expression. Increased copy number has been suggested to act as a defence mechanism against the damage caused by ROS production, which may in turn be a consequence of defective mitochondria [254, 275]. There is evidence that increased copy number can be maladaptive [273]; however it remains to be determined whether the response seen here is a secondary adaptation or part of the primary pathological process.

In order to pursue this mechanism further, there are three main strategies. First, we need to sequence a large number of similarly affected individuals to see whether this polymorphism occurs in any other families with renal disease. I have already sequenced 46 carefully selected individuals; however such a study would ideally consider hundreds of individuals. The rarity of this haplogroup in the UK makes it hard to select an ideal comparator population. It would seem logical to concentrate on the population geographically close to this family, although this might just identify other individuals descended from a common founder.

Second, given the limitations of culture in high glucose medium discussed, it might be informative to see how these cells behaved in a medium where glycolysis was inhibited. One possibility here would be to culture the cells in galactose which reduces the amount of ATP generated by glycolysis [254, 285]. Third, a significant limitation of these *in vitro* studies is the inability to control for nuclear background. This could be circumvented by the generation of trans-mitochondrial cybrids, which would enable direct comparison of variant and wild-type mtDNA sequences. A further advantage of this experiment is that it is possible to use fresh platelets as mitochondrial donors, thereby avoiding the period of time in culture required in these fibroblast experiments.

Chapter 6 Familial erythrocytosis and von Hippel-Lindau disease

6.1 Introduction

6.1.01 Von Hippel-Lindau disease

One of the best characterised genetic conditions affecting the kidneys is von Hippel-Lindau (VHL) disease. This is a dominantly inherited cancer syndrome (MIM 193300) caused by mutations in the *VHL* tumour suppressor gene. More recently it has become apparent that mutations in this gene are also responsible for cases of congenital polycythaemia (also known as familial erythrocytosis-2; MIM 263400), as well as playing a role in pulmonary vascular tone [286]. In this chapter, the role of the *VHL* gene was studied in detail when a young boy presented to clinical services with life-threatening pulmonary hypertension and erythrocytosis and was found to be a compound heterozygote for two novel *VHL* mutations [287].

VHL is a highly conserved gene that gives rise to a ubiquitously expressed mRNA. It forms an important component of a molecular pathway used by metazoans to sense and respond to changes in oxygen [104]. The *VHL* gene encodes two proteins as a result of alternative start codons [288-290] (Figure 6.1). These two proteins migrate with an apparent molecular weight of 24-30 kD and 19 kD and have overlapping biological functions [104]. The VHL protein contains two functional sub-domains, termed alpha and beta [291]. The alpha domain (residues 155-192) binds to elongin C, and forms part of a multiprotein E3 ubiquitin ligase complex [292, 293]. The beta domain (residues 63-154) consists of a largely hydrophobic region which acts as a substrate recognition site [291].

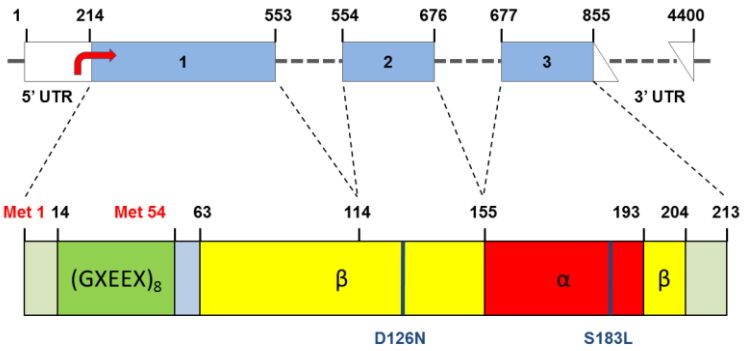


Figure 6.1 Structure of VHL

The location of the two mutations studied in this chapter (D126N and S183L) are indicated. The alternative start codons are marked by the methionine (Met) residues; the 19 kDa isoform contains amino acids 54-213 of the full-length protein.

One of the best characterised roles of VHL is its ability to regulate the alpha subunit of HIF [125]. The HIF pathway has attracted considerable attention in cancer biology because HIF acts as a master transcriptional activator of over a hundred genes that generally promote adaptation to low oxygen conditions (hypoxia). HIF is a highly conserved, sequence-specific DNA-binding basic helix-loop-helix transcription factor. The active form is a heterodimeric protein composed of a constitutively expressed beta subunit and one of three alpha subunits (HIF-1α, HIF-2α, and HIF-3α). HIF activity is regulated by post-translational modification of the alpha subunit, collectively termed HIF-α.

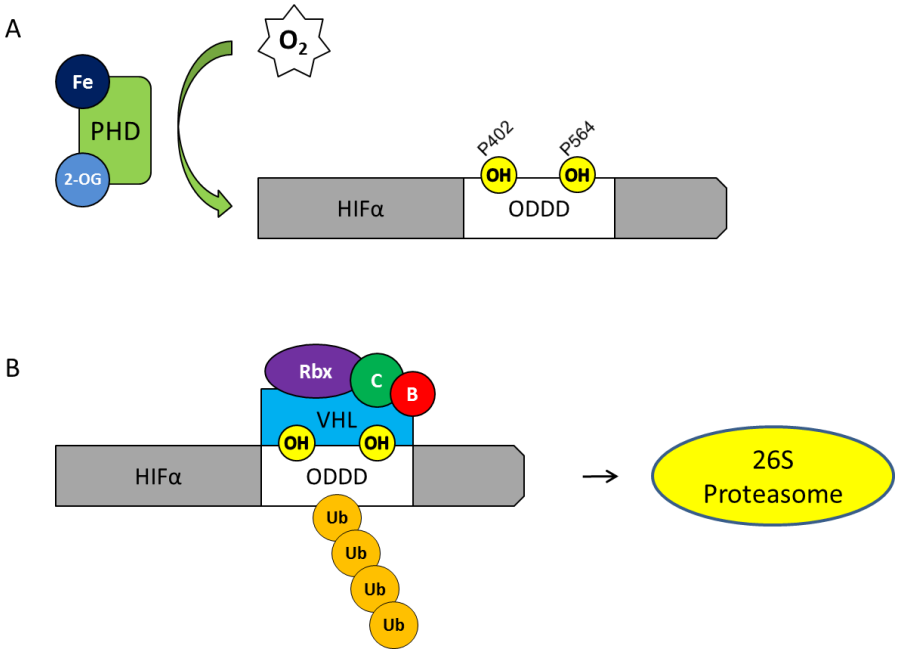


Figure 6.2 Regulation of HIFα by VHL

A. HIF-α is hydroxylated by the prolyl hydroxylase domain-containing proteins (PHD) at two specific proline residues (P402 and P564 in HIF-1α) within the oxygen-dependent degradation (ODD) domain. This reaction uses molecular oxygen as a substrate, and the co-factors iron (Fe) and 2-oxoglutarate (2-OG). B. The hydroxylated form of HIF-α is then recognised by VHL, which acts as an E3 ubiquitin ligase (in complex with elongin B, elongin C, and ring box 1 (RBX)) and mediates the ubiquitination of lysines within the ODD domain. The ubiquitinated form of HIF-α is then targeted for degradation within minutes by the 26S proteasome.

The interaction between VHL and HIF-α is critically dependent on the levels of oxygen (Figure 6.2). VHL only recognises HIF-a subunits that have been hydroxylated on two specific proline residues, at positions 402 and 564 in HIF-1α [294-297]. This hydroxylation reaction requires molecular oxygen and 2-oxoglutarate, and is carried out by one of three prolyl hydroxylase (PHD) enzymes [298, 299]. Following binding to VHL, HIF-1α is ubiquitinated on three

conserved lysines within the oxygen dependent degradation domain [300]. It is then targeted for destruction by the 26S proteasome [301].

When oxygen levels are low, or when VHL function is compromised, HIFa subunits rapidly accumulate and form an active transcriptional complex. The active transcription factor directly activates genes containing hypoxia response elements (HREs), as well as interacting with other transcriptional control complexes, and influencing the expression of other transcription factors and signalling pathways. These genes include energy metabolism, angiogenic signalling, vasomotor regulation, cell growth and apoptosis.

VHL also has a large number of HIF-independent functions that may help to explain why loss of function is associated both with tumour formation and with erythrocytosis [302]. These functions include regulation of apoptosis and transcription [303-305], and association with the extra-cellular matrix and microtubule cytoskeleton [306-309].

6.1.02 Erythrocytosis and the oxygen-sensing pathway

Erythrocytes are the vertebrate organism's principal means of delivering oxygen to the body tissues. Both erythrocyte number and oxygen carrying capacity are able to respond to fluctuations in oxygen tension. Control over erythrocyte number is mediated by the glycoprotein hormone erythropoietin (EPO). EPO levels are determined primarily at the level of gene transcription, by HIF binding to an HRE in the *EPO* promoter [310]. As detailed above, HIF levels are determined by the availability of molecular oxygen. Recently, mutations in *VHL*, *PHD2*, and *HIF2A* have been associated with congenital erythrocytosis, providing a clear link between the cellular oxygen sensing pathway and systemic responses to hypoxia.

Erythrocytosis is defined as an absolute increase in the red cell mass. It can result from defects at any stage in the generation of erythrocytes, which can be divided into four broad categories (Table 6-1). The first familial erythrocytosis to be characterised was due to mutations of the EPO receptor (MIM 133100). Mutations in the EPO receptor result in receptor hypersensitivity and are associated with subnormal or undetectable levels of EPO [311]. The other three recognised forms of familial erythrocytosis (MIM 263400, 609820, 611783) are due to mutations in the oxygen sensing pathway, and are associated with elevated or inappropriately normal levels of EPO.

Aetiology	Congenital / Acquired	Primary / Secondary
Defect in erythroid precursors	Acquired	Primary
Tissue hypoxia / EPO-secreting tumour	Acquired	Secondary
Impaired oxygen delivery by haemoglobin	Congenital	Secondary
Defect in oxygen sensing pathway	Congenital	Secondary

Table 6-1 Clinical basis of erythrocytosis

The first erythrocytosis-associated mutation in the oxygen sensing pathway was reported in a large cohort of patients from the Chuvash region of the former Soviet Union, in association with an autosomal recessive hereditary erythrocytosis syndrome originally described by Polyakova in 1974 [312]. This condition was initially thought to be endemic to the Chuvash Republic, with a frequency of approximately 1 in 10,000 amongst a population of 1 million. It has subsequently been found in patients from Ischia, North America, Germany, Turkey, and Bangladesh [313-316]. The mutation responsible is a 598C>T transition in exon 3 of the *VHL* gene. This mutation results in an amino acid change of arginine to tryptophan at residue 200 in the VHL protein (R200W).

The consequences of the R200W *VHL* mutation at the cellular level have been studied in transformed B lymphocytes derived from patients homozygous for this mutation [317, 318]. This mutation does not affect the level of VHL protein stability. However, the mutant protein showed decreased affinity for, and reduced ubiquitination of, HIF-1α. This was associated with increased HIF-1α protein under normoxic conditions and comparatively normal levels in hypoxia. Increased HIF-1α protein levels were associated with increased mRNA expression of a subset of known HIF target genes, including *EPO*, vascular endothelial growth factor (*VEGF*), and glucose transporter 1 (*GLUT1*) [317, 318]. Perrotta *et al.* reported an increase in normoxic levels of *VEGF* but not *EPO* mRNA in transformed B lymphocytes carrying the R200W mutation [318].

Taken together, these findings suggest that the R200W mutation in *VHL* impairs VHL-mediated degradation of the HIF-1α protein under normoxic conditions. The balance of the HIF transcription system is therefore altered, resulting in increased transcription of at least some HIF-regulated genes. Up-regulation of these HIF target genes is therefore hypothesised to explain the observed phenotype. The prevalence and magnitude of altered gene transcription

of HIF-regulated genes in different cell types as a result of this mutation have still not been clearly established.

In addition to the R200W mutation, other *VHL* gene mutations have also been reported to be associated with congenital polycythaemia [313, 315, 316]. Such cases are rare, but support the idea that altered VHL function is a significant cause of congenital polycythaemia with elevated EPO. These mutations are distinct from those associated with classical VHL disease. They lie at the C-terminal end of the protein and are hypothesised to result in a more modest loss of function [119]. There is also evidence to suggest these C-terminal mutations exhibit a heterozygote phenotype, both with regard to HIF-1α stabilisation [318], and in measurements of physiological parameters such as haemoglobin or blood pressure [319-321].

An important question is why these C-terminal *VHL* mutations cause erythrocytosis without the classical tumour syndrome. One possibility is that the mutant proteins have insufficient effect on HIF-α levels, with initiation of CCRCC and haemangioblastoma requiring a greater degree of HIF activation. Alternatively, these mutations may preserve some specific function of VHL other than the ubiquitination of HIF-α. Note that vertebral haemangiomas have been observed in Chuvash erythrocytosis, but it is not known whether they display loss of heterozygosity or whether this may be a reflection of the increased levels of VEGF observed in these patients [129].

In addition to erythrocytosis caused by mutations in *VHL*, there are now two other types of familial erythrocytosis that are caused by mutations in the oxygen-sensing pathway, affecting *PHD2* and *HIF2A*. By contrast with VHL-associated erythrocytosis, these mutations behave in an autosomal dominant fashion. However, they all suggest that erythrocytosis can result from abnormal activation of this pathway.

The first *PHD2* mutation to be described in association with erythrocytosis, P317R, impairs binding to HIF-α, diminishes the capacity of PHD2 to down-regulate an HRE reporter gene, and results in a loss of function against HIF-α [322]. Subsequent mutations have also been shown to have substantial effects on the catalytic domain of the protein, whether local missense mutations or frame shifts which are N-terminal to it [323-325]. This has two implications: first, that erythrocytosis is caused by haploinsufficiency, and second, that PHD2 is the critical PHD isoform controlling EPO in humans. This in turn implies that PHD2 loss of function is sufficient to allow HIF to escape PHD2 and PHD3 negative feedback suppression in EPO-producing cells

[286]. This notion is supported by a number of mouse knockout models that have demonstrated a critical role for PHD2 in the control of red cell mass [326, 327].

In contrast to loss of PHD2 function, mutations in *HIF2A* causing erythrocytosis have been shown to cause partial gain of function. Thus the G537W mutant impairs PHD2-mediated hydroxylation and subsequent recognition of the hydroxylated HIF-2α by VHL, resulting in normoxic stabilisation of HIF-2α [128]. All mutations in *HIF2A* described thus far are concentrated in a small region located C-terminal to the primary hydroxyl-acceptor proline (residue 531) [127, 328]. There is heterogeneity in the functional defects associated with *HIF2A* mutations; whereas many mutations impair interaction with both PHD2 and VHL, the M535V mutation only affects interaction with PHD2 [329]. These findings would suggest that HIF-2α has a critical role in EPO control in humans, a position that is supported by both loss-of-function and gain-of-function mouse models [330, 331]. However, HIF-2α may have a role in the control of red cell mass beyond the regulation of EPO, as seen by HIF-2α regulation of vascular cell adhesion molecule-1 (VCAM-1) in haematopoietic tissue and its role in promoting intestinal absorption of iron [286].

6.1.03 Pulmonary hypertension and the oxygen sensing pathway

Much less is known about how mutations in the oxygen sensing pathway are able to cause alterations in cardiac and pulmonary responsiveness. Pulmonary hypertension has been observed in association with erythrocytosis-causing mutations of *VHL* and *HIF2A* but not *PHD2* [127, 319, 332]. Patients with Chuvash polycythaemia exhibit disordered vascular physiology, showing a higher incidence of vertebral haemangiomas, varicose veins, lower blood pressures, and elevated serum VEGF concentrations [129]. These patients show an exaggerated physiological response to hypoxia, characterised by abnormalities in respiratory and pulmonary vascular regulation [333, 334]. This response is similar to that seen in association with acclimatisation to the hypoxia of altitude. Studies of the Chuvash knock-in mouse support these findings, with evidence of pulmonary vasoconstriction and enhanced normoxic ventilation, similar to that seen in human patients with this mutation [335]. Superimposing decreased HIF-2α activity resulted in partial protection against pulmonary vascular remodelling, haemorrhage and oedema [335].

6.2 Results

6.2.01 Clinical Background to the D126N and S183L mutations in VHL

In 2011, Bond *et al.* reported a novel compound heterozygous mutations of *VHL* in a boy presenting at two months of age with a severe clinical phenotype [287]. He presented with severe pulmonary artery hypertension, and required medical treatment in an intensive care unit. Consistently raised haemoglobin concentrations of over 21 g/dL (normal range 11-16 g/dl) prompted further investigation. Serum erythropoietin concentration was grossly elevated at 4120 IU/L (normal range 2.5-10), greatly in excess of that expected with chronic hypoxia. Apart from EPO levels, serum levels of HIF target genes, such as *VEGF* and plasminogen activator inhibitor-1 (*PAI-1*) [319], were not elevated. Regular venesection was commenced and coincided with a steady amelioration of his cardiovascular symptoms and pulmonary vascular measurements. Now nearly eleven years old, his clinical course has far exceeded initial prognostic predictions.

Given the combination of erythrocytosis with elevated EPO levels, and early onset suggesting a congenital condition, genes from the oxygen sensing pathway were sequenced. The patient was found to be a compound heterozygote for mutations in exon 2 (376 G>A) and exon 3 (548 C>T) of the *VHL* gene, coding for the amino acid changes Asp126Asn (D126N) and Ser183Leu (S183L) respectively (Figure 6.1). Sequencing of DNA obtained from the patient's parents revealed heterozygosity for 376 G>A in the mother and 548 C>T in the father.

His parents are genetically unrelated and come from distinct ethnic populations (Caucasian and North African). They appear clinically unaffected and have normal haemoglobin concentrations. They have no evidence of pulmonary hypertension or features of VHL disease. He has two siblings, neither of whom carries either mutation.

6.2.02 Expression of VHL mutants in a renal cancer cell line

In order to study the effect of the D126N and S183L mutations identified in this boy, a VHL-deficient renal cancer cell line was first used to study the effect of these mutations separately. The mutations found in this patient were compared with four controls: the Chuvash polycythaemia mutant, R200W, a ß-domain mutant, N78S, which is deficient in HIF-binding, a wild-type VHL positive control and an empty vector negative control. The RCC10 cell line has a deletion of nucleotides 474-477 in *VHL*, resulting in a premature stop codon and no detectable

VHL mRNA [162]. These cells therefore allow the reconstitution of wild-type or mutant VHL function in the absence of endogenous VHL function. Figure 6.3 shows that levels of both HIF-1α and HIF-2α protein are increased in cells expressing the D126N and S183L *VHL* mutants when grown at ambient oxygen levels.

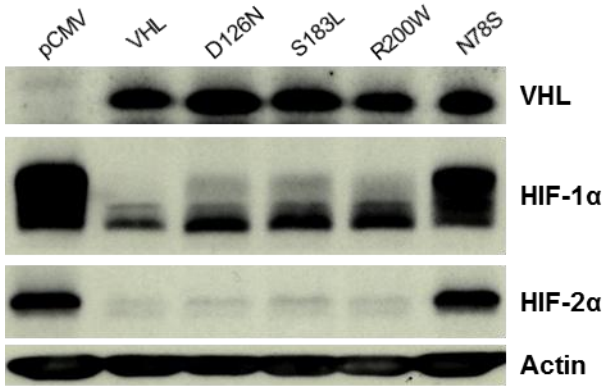


Figure 6.3 The D126N and S183L VHL mutations result in increased basal levels of HIF-α protein

RCC10 cells expressing the D126N or S183L mutants show increased levels of HIF-1α and HIF-2α protein compared with cells expressing the wild-type protein (VHL). This is comparable to the R200W mutant and less than the N78S mutant that is deficient in HIF binding and the empty vector control (pCMV).

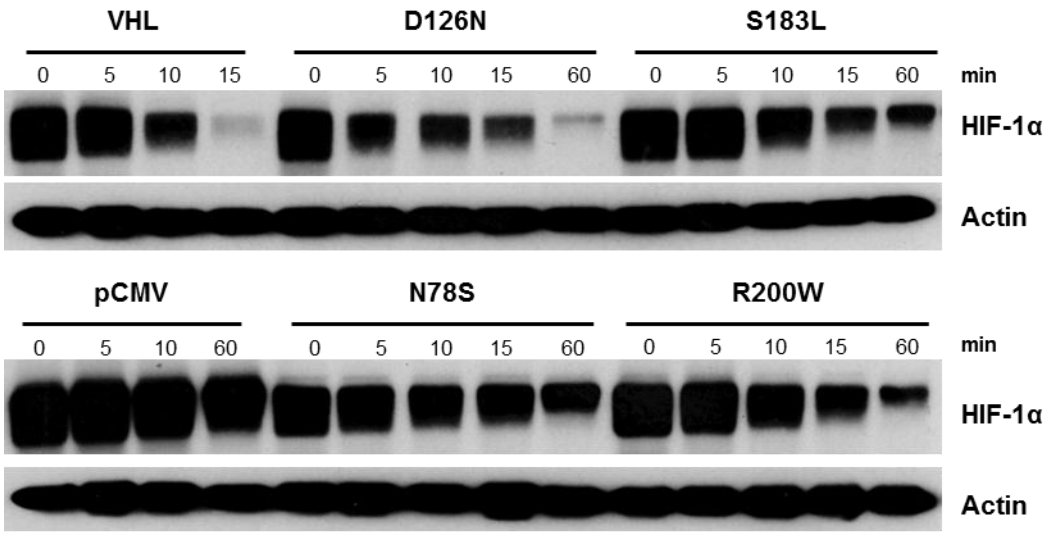


Figure 6.4 The D126N and S183L VHL mutations result in impaired degradation of HIF-1α

The rate of reduction in HIF-1α protein levels was measured following addition of cycloheximide to inhibit protein translation. The half-life of HIF-1α was increased in RCC10 cells expressing the D126N and S183L mutants, comparable to the R200W mutant and less than the N78S mutant.

In order to show that the increase in HIF-1α observed in Figure 6.3 might be due to reduced protein destruction, I studied the degradation of HIF-1α. Following stabilisation of HIF-1α by exposure to 1% O2 for 16 hours, the cells were treated with cycloheximide to inhibit protein synthesis and then re-exposed to oxygen. Figure 6.4 shows that both mutants result in reduced HIF-1α degradation compared to wild-type, in a manner analogous to the Chuvash mutant. This could be due to reduced HIF-1α binding, impaired ubiquitin ligase activity, or both.

6.2.03 Expression of VHL in patient-derived LCL cells

In an attempt to study the biochemistry of these mutant proteins in patient samples I obtained a population of B-cells (LCL cells) that had been immortalised by infection with Epstein-Barr virus (EBV) from the patient described above and healthy controls. This procedure was carried out on two separate occasions, using different blood samples, and different donors for the control lines. Exons 2 and 3 of *VHL* were sequenced in all LCL lines used to confirm their genotype (Figure 6.5).

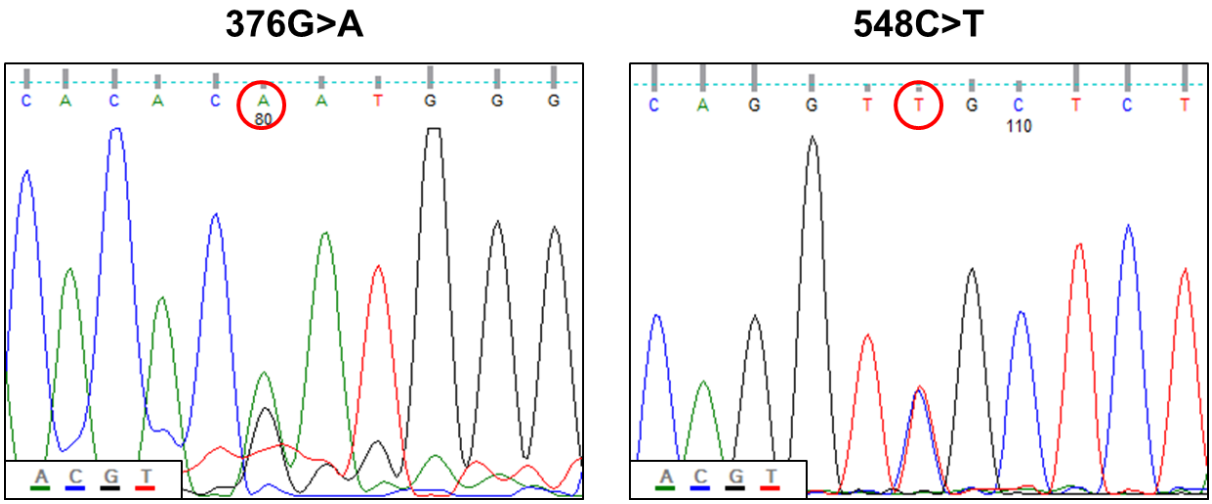


Figure 6.5 VHL sequence in patient-derived LCL cells

Heterozygosity for the D126N and S183L mutations was confirmed by Sanger sequencing in both mutant LCL lines. Control lines were confirmed to have the wild-type sequence.

As shown in Figure 6.6, the D126N and S183L mutations did not affect the level of *VHL* mRNA in LCL cells. However VHL protein expression was reduced in the LCL cells derived from the patient (Figure 6.6). In common with a number of other cell lines, LCL cells predominantly express the 19kD isoform of VHL. Previous studies would suggest that erythrocytosis is caused by impaired degradation of HIF-α under conditions of ambient oxygen. However there was no increase in the levels of HIF-1α in normoxia (Figure 6.6). LCL cells do not express HIF-2α so this could not be assessed [336].

Given the evidence of abnormal HIF-1α degradation in RCC10 cells, I sought to replicate this experiment in LCL cells. Following stabilisation of HIF-1α by exposure to 1% O2 for 16 hours, the cells were treated with cycloheximide to inhibit protein synthesis and then re-exposed to oxygen. Figure 6.7 shows that the mutant cells exhibit reduced HIF-1α degradation compared

to wild-type. This suggests that there is impaired degradation of HIF-1a even when both mutants are expressed at physiological levels.

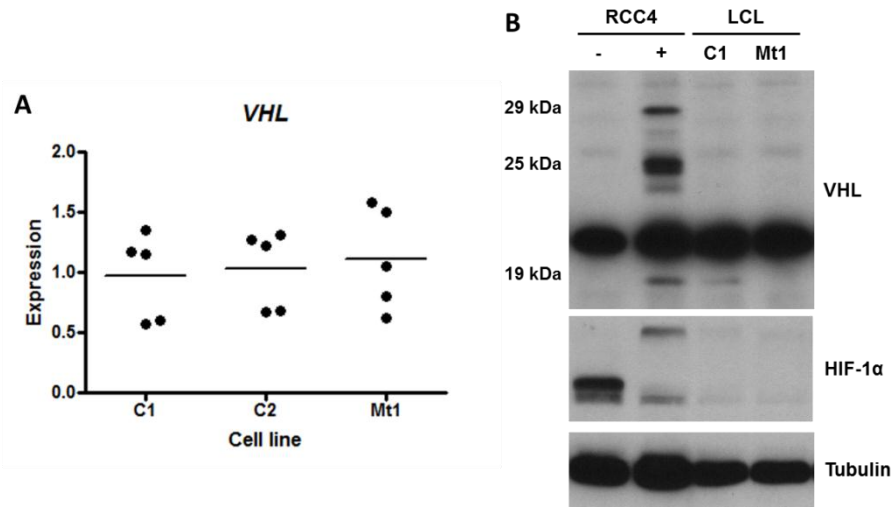


Figure 6.6 Expression of VHL in lymphoblastoid cells

A. expression of VHL mRNA is equivalent in control (C1 and C2) and mutant (Mt1) cell lines. B. VHL protein expression in lymphoblastoid cells compared to a renal cancer cell line (RCC4) that lacks VHL (-) or with full-length VHL reconstituted (+). Due to post-translational modification a number of species are recognised by the VHL antibody; the 19 kDa isoform is the main one in LCL cells. Despite reduced VHL expression, there is no increase in basal levels of HIF-1α.

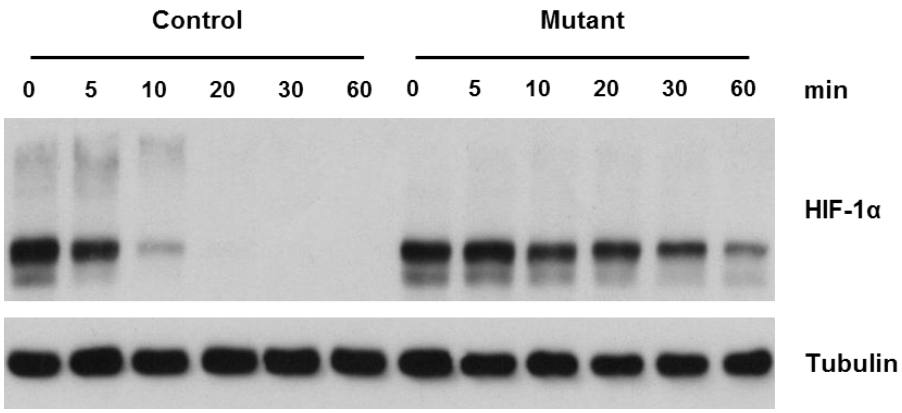


Figure 6.7 Degradation of HIF-1α in lymphoblastoid cells

The rate of reduction of HIF-1α protein levels was measured following treatment with cycloheximide to inhibit protein translation and exposure to oxygen for the times indicated.

6.2.04 Abnormal ubiquitination of HIF-1α in hypoxia

Both mutant and control LCL cells demonstrated increased levels of HIF-1α in hypoxia, due to the decreased availability of oxygen for hydroxylation (Figure 6.8). Treatment with dimethyloxalylglycine (DMOG) or desferrioxamine (DFO) also led to increased levels of HIF-1α. DMOG is a 2-oxoglutarate analogue that non-specifically inhibits PHD activity, while DFO is a more potent agent that prevents HIF hydroxylation by chelation of the ferrous ion required as a co-factor by the PHD enzymes for the hydroxylation reaction. Treatment with either of these agents prevents hydroxylation of HIF-a, thus leading to stabilisation of the protein.

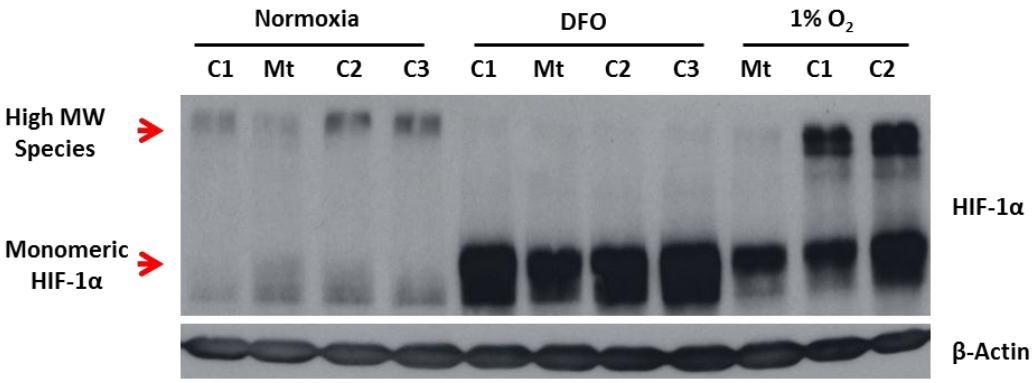


Figure 6.8 Mutant LCL cells show reduced levels of high MW HIF species

Treatment of LCL cells with desferrioxamine (DFO) blocks prolyl hydroxylation and results in the stabilisation of unmodified HIF-1α. Treatment with hypoxia results in stabilisation of HIF-1α. Mutant LCL cells show reduced levels of high MW species, which may correspond to poly-ubiquitinated HIF-1α.

In addition to a band corresponding to monomeric HIF-1α at around 120kD, all the control lines exhibited a number of higher molecular weight species when exposed to 1% oxygen (hypoxia), but not when treated with DFO or DMOG (Figure 6.8). I hypothesised that these higher molecular weight species detected may correspond to poly-ubiquitinated HIF-1a molecules. Importantly, these were consistently absent from the mutant cell line (Figure 6.8). This suggests that the VHL mutant LCL cells may have defective ubiquitin ligase activity.

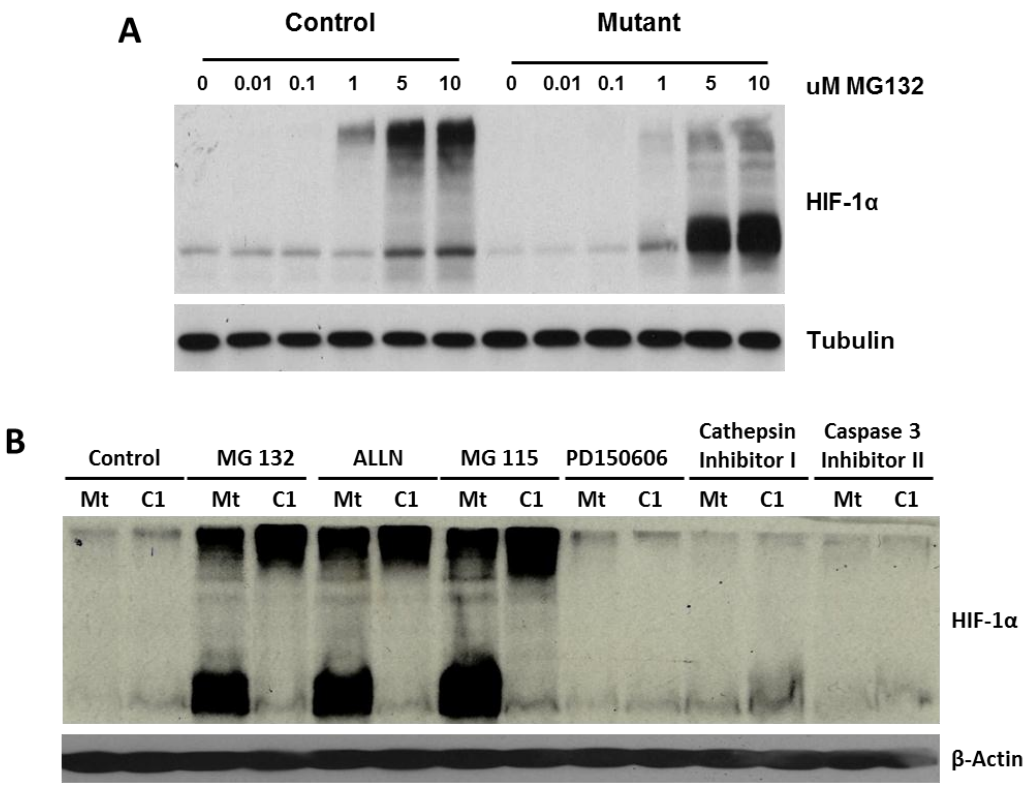


Figure 6.9 Mutant LCL cells are deficient in HIF-1α ubiquitination

Accumulation of poly-ubiquitinated HIF-1α is seen in control but not mutant LCL cells (A) and is specific to treatment with proteasomal inhibitors (B).

Addition evidence showing that mutant LCL cells are deficient in HIF-1α ubiquitination was provided by pharmacological inhibition of proteasomal degradation. MG132, MG115, and ALLN are all peptide aldehyde inhibitors of the proteasome with respectively decreasing specificity. Inhibition of the proteasome leads to accumulation of ubiquitinated HIF-1a in control, but not mutant cells (Figure 6.9). It is important to note that all of these inhibitors may have off-target effects against calpains and cathepsins. To control for this, LCL cells were also treated with PD150606, which is a more specific calpain inhibitor, caspase inhibitor II and cathepsin inhibitor I, which are more specific to their respective peptidases. Figure 6.9 shows that accumulation of high molecular weight HIF-1α species was only seen in LCL cells treated with inhibitors of the proteasome. This implies that the degradation of HIF-1α in both mutant and control LCL cells is proteasome dependent.

6.2.05 Expression of endogenous HIF-1α target genes

Given the evidence for abnormal ubiquitination and impaired degradation of HIF-1α, I examined the expression of various HIF target genes in LCL cells (Figure 6.10). *VEGF* is a well-characterised HIF target gene that has been used in previous studies of familial erythrocytosis [317, 318]. Prolyl-4 hydroxylase alpha-1 (*P4HA1*) and 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4 (*PFKFB4*) are both HIF target genes that show a particularly high level of hypoxic induction in LCL cells [130].

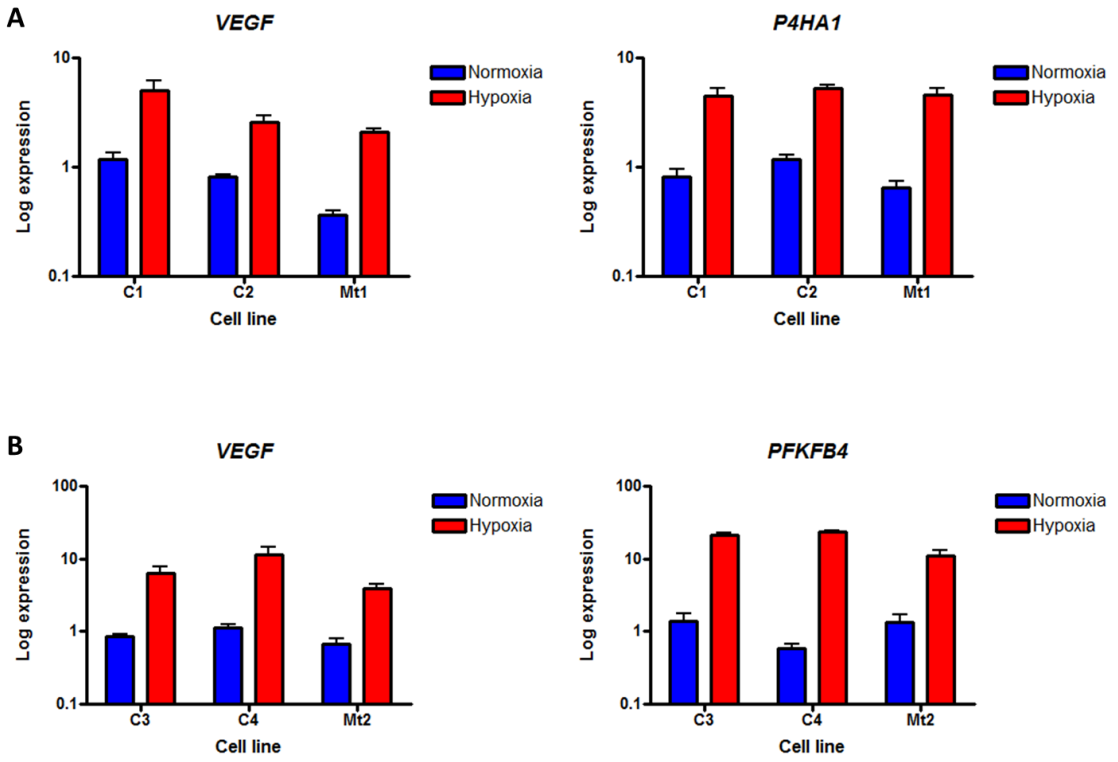


Figure 6.10 Expression of HIF target genes is not affected by the D126N and S183L VHL mutations

The experiment was repeated for independent transformations (A and B) of B-cells from the patient (Mt1 and Mt2) and controls (C1-C4).

Gene expression was assessed at baseline (normoxia) and after induction by overnight culture at 1% O_2 (hypoxia). All genes tested showed highly significant induction by hypoxia ($p < 0.0001$). Surprisingly, *VEGF* showed significantly reduced expression in normoxia ($p < 0.005$; Figure 6.10A), however this difference could not be replicated (Figure 6.10B). I also examined a number of other HIF target genes, including *erythropoietin, HIF-2α, PHD3, adrenomedullin, aldolase C*, with similar results (data not shown).

Previous work on the Chuvash polycythaemia mutant in LCL cells had only shown a difference in baseline function [317, 318]. These authors did not show any difference in hypoxic induction of gene expression Furthermore, the normal inter-individual differences in HIF pathway gene expression are seen to diminish at extreme hypoxia conditions [335]. Given that no difference in gene induction was observed at more permissive oxygen levels, it is not surprising that there was no difference in hypoxia.

6.2.06 VHL knockdown

In addition to reconstituting individual *VHL* mutants in a *VHL*-deficient cell line, I also looked at the effect of *VHL* knockdown in LCL cells. LCL cells typically grow in suspension and I therefore used lentivirus-mediated introduction of a short hairpin RNA (shRNA) targeting *VHL*. This approach achieved 70-80% knockdown in a pool of LCL cells, as measured by RT-PCR (Figure 6.11), and a marked reduction in VHL protein levels (Figure 6.11). These cells showed increased levels of HIF-1α in normoxia in both mutant and control lines (Figure 6.12). Moreover, they also showed a decrease in the level of higher molecular weight HIF-1α species in both control LCL cells.

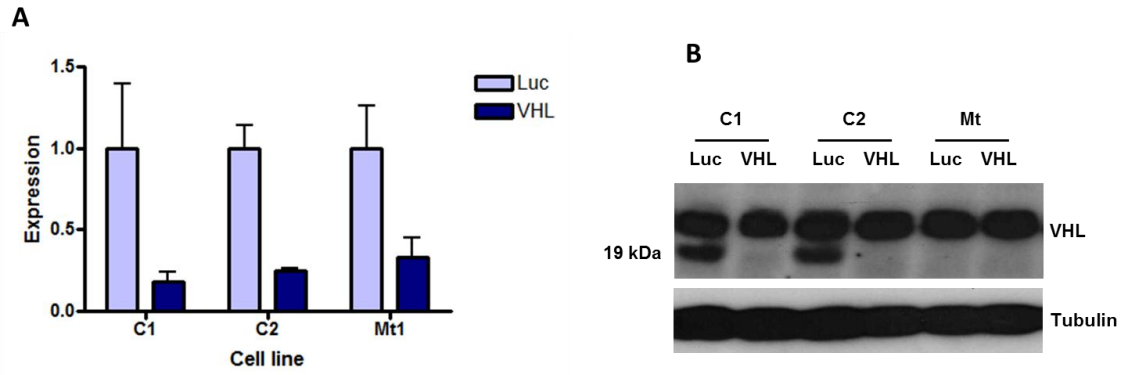


Figure 6.11 Lentiviral-mediated shRNA infection results in efficient knockdown of VHL mRNA and protein

VHL expression was analysed in LCL cells treated with VHL shRNA (VHL) and control (luciferase, Luc) by RT-PCR (A) or immunoblotting (B).

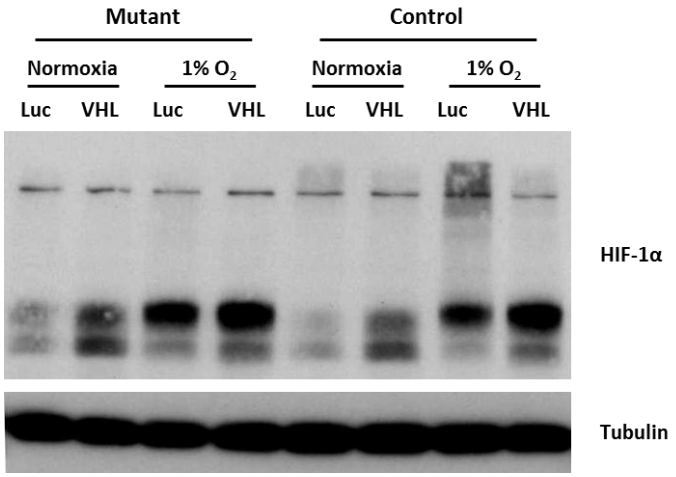


Figure 6.12 VHL knockdown increases HIF-1α levels in normoxia LCL cells in both normoxia and hypoxia

VHL knockdown also decreases poly-ubiquitinated HIF-1 α in control cells in hypoxia.

6.2.07 Response to proteasomal inhibition following VHL knockdown

Treatment of control cells with *VHL* knockdown (Figure 6.13A) shows a remarkable recapitulation of the failure to accumulate poly-ubiquitinated HIF-1α seen in mutant cells (Figure 6.8). Treatment of mutant cells with *VHL* knockdown does not show additional reduction of poly-ubiquitinated HIF-1a (Figure 6.13B). This provides further evidence that the higher molecular weight HIF-1α species are related to VHL ubiquitin ligase activity, which is reduced in both VHL knockdown and the D126N and S183L mutants.

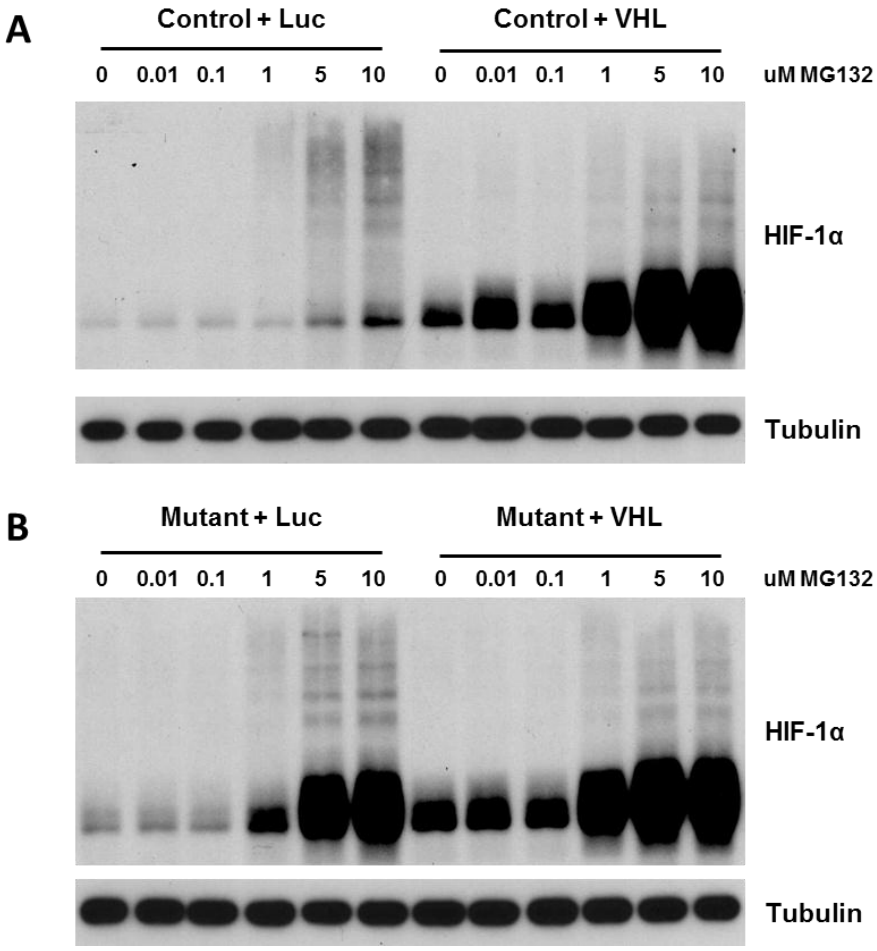


Figure 6.13 VHL knockdown impairs ubiquitination of HIF-1α

Control (A) or mutant (B) LCL cells were treated with a lentivirus encoding control (Luc) or VHL-specific (VHL) shRNA. Subsequent treatment with MG132 at the concentrations indicated showed that VHL knockdown decreases poly-ubiquitinated HIF-1a in control cells in a similar manner to that seen in mutant cells.

6.2.08 Expression of endogenous HIF-1α target genes following VHL knockdown

Given that VHL knockdown was sufficient to increase levels of HIF-1α protein in normoxia, the expression of various HIF target genes in LCL cells was then examined (Figure 6.14). Surprisingly, while hypoxia resulted in significant induction of HIF target genes, VHL knockdown did not increase levels of HIF target genes. *VEGF* was the only gene that showed any change in expression following VHL knockdown, and this was only apparent in the mutant cells. This might indicate that HIF is only able to activate target genes once VHL activity has been reduced below a certain threshold.

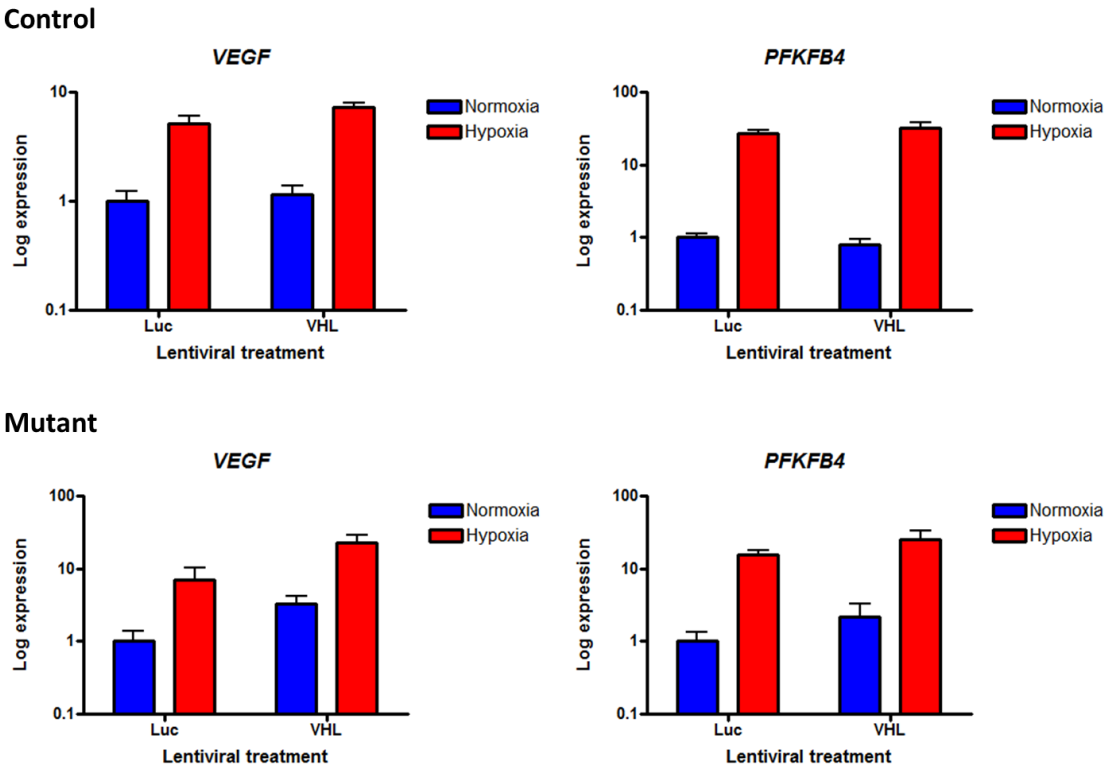


Figure 6.14 Expression of HIF target genes in control and mutant LCL cells treated with VHL knockdown

VEGF expression was increased in mutant LCL cells treated with VHL knockdown at baseline and in hypoxia ($p < 0.05$), but there was no significant difference in the expression of PFKFB4.

6.3 Interpretation

6.3.01 Justification for the use of a transformed B lymphocyte model to assess the HIF-system

A number of previous studies of erythrocytosis caused by mutations in the oxygen sensing pathway have employed transformed B lymphocytes [130, 317, 318, 322, 333]. Using such cells ex-vivo offers the chance to conduct extreme physiological manipulation of the HIF pathway in human cells. Such cells retain the genetic identity of their donor, and thereby allow the detailed examination of rare human mutations.

EBV is the only virus known to immortalise human B lymphocytes. Only a proportion of the circulating B cells are immortalised (~1 in 100), and resting cells are immortalised in preference to activated ones [337]. T cells from EBV seropositive individuals suppress B cell immortalisation by EBV in culture, hence the need to remove these cells by selection. EBV-immortalised B cell lines are initially polyclonal and secrete all major classes of immunoglobulin, although after prolonged culture *in vitro* they become oligoclonal or monoclonal [338]. LCL cells can be cultured for 150-200 passages, at which point their characteristics change [339]. A minority will senesce, while the majority develop chromosomal instability and become immortalised. Although early passage cells were used in these experiments, it is important to note that it takes four weeks to generate sufficient cells to conduct experiments, and a further four weeks following lentiviral infection.

It is clear that the process of immortalisation has significant effects on cellular function, and that there are significant differences between the function of independent transformations of B cells from the same donor [130]. Alternative strategies have employed non-transformed peripheral mononuclear cells (PBMCs), or a number of separate EBV transformations [130]. I was able to study two independent transformations of B-cell from this patient, giving largely consistent data.

6.3.02 Reduced expression of VHL

The D126N and S183L *VHL* mutations provide an ideal opportunity to dissect the molecular pathophysiology of this patient. The oxygen-sensing pathway is well understood, but this patient would be the youngest to present with erythrocytosis and pulmonary hypertension. This could imply that these mutations have a more profound effect than those previously

studied. Our collaborators and I therefore looked first at the effect of a introducing the individual mutants into a VHL-deficient renal cancer cell line (RCC10).

The D126N and S183L mutants showed reduced protein stability compared to wild-type following cycloheximide treatment [287]. VHL is degraded by the proteasome, and it is known that both elongin B and C are required for maximum stabilisation [340]. Thus the S183L mutation, which is predicted to lie within the alpha-domain at the surface, could impact on binding to other components of the ubiquitin ligase complex. These data provide evidence that both mutations contribute to the patient's phenotype, but do not differentiate between a loss of function and a dominant negative mechanism of action.

Impairment of VHL function can lead to the sustained rates of high rate of anaerobic-like glycolysis even in the presence of oxygen, known as the Warburg effect [341, 342]. This has been shown to result from HIF-mediated enhancement of glycolysis and suppression of mitochondrial respiration [343]. Our collaborators therefore assessed VHL protein function *in vitro* by measurement of the pH of the media used to culture RCC10 cells carrying the D126N and S183L mutants [287]. Both these mutants exhibited an effect on pH that was intermediate between wild-type and the N78S mutant [287].

I examined whether the Warburg effect was associated with HIF-α stabilisation in these cells. Both mutants showed stabilisation of HIF-1α and HIF-2α that was again intermediate between wild-type and the N78S mutant and similar to the levels seen with the Chuvash mutation. In order to examine this in more detail I looked at the effect of *VHL* mutation on HIF-1α stability. Previous experiments have shown that the half-life of HIF-1α is just a few minutes, extending to over an hour in cells deficient for VHL [125]. Our results confirm that both mutants impair HIF-1α degradation, with kinetics comparable to that seen with the R200W mutant.

Since stably transfected cell pools exhibit a range of expression of the introduced protein I next looked at a more physiological model, using patient-derived lymphoblastoid cells. This allowed me to look at cells with a normal level of VHL expression, but meant that I had to look at the function of both mutations together. Although there are now a number of reported mutations in *VHL* associated with erythrocytosis [119], only a minority of these have looked at functional studies in lymphoblastoid cells. I found that these cells showed normal levels of *VHL* mRNA, which is in agreement with studies showing that the Y175C mutant did not affect *VHL* mRNA levels [320]. Importantly, however I saw a marked reduction of VHL protein expression. This

may well be explained by the reduction in VHL protein stability seen in RCC10 cells, but it stands in contrast to previous studies [317].

VHL exists in two isoforms with overlapping function [104]. In order to confirm the specificity of the p19 species that we detected most abundantly in LCL cells, we used a lentiviral mediated shRNA to knockdown VHL in these cells. Knockdown of VHL leads to a reduction in protein levels comparable to that seen in the mutant LCL cells. This was associated with an increase in HIF-1α stabilisation in normoxia in both control and mutant LCL cells, a finding that has been reproduced in other cell lines [344]. This shows that these mutants cause a loss of function, rather than acting in a dominant negative manner.

6.3.03 Functional effects

Given the impaired Warburg effect seen in RCC-10 cells [287], and the reduced levels of VHL expression in LCL cells (Figure 6.6), I expected that the LCL cells would show evidence of activation of the oxygen sensing pathway. HIF-1a is a highly labile protein, and consequently small differences in protein levels may be affected by experimental technique. I therefore looked for more robust evidence of HIF activation, manifest as increased expression of HIF target genes.

Despite the wide variety of genes that are modified by HIF levels, previous studies of erythrocytosis have concentrated on a few well-established HIF-target genes, such as *VEGF* or *EPO*. Given the importance of cell type in determining the abundance of gene expression, a more robust approach is to examine those genes which show the greatest hypoxic response specifically in B cells or their EBV-transformed derivatives. *VEGF*, *P4HA1,* and *PFKFB4* have all been shown to have significant up-regulation in response to hypoxia in LCL cells, and have been used before to identify inter-individual differences in HIF pathway biology [130]. Analysis of expression in normoxia provides the most sensitive test for alterations in function, since inter-individual differences are diminished in hypoxia [333]. It is also important for semi-quantitative RT PCR to ensure that the reference gene transcript is not affected by hypoxia.

It remains intriguing to consider how activation of different HIF pathways might play a particular role in the pathophysiology of these patients. Clearly, the activity of HIF on the *EPO* promoter must be of major relevance in the generation of erythrocytosis. That does not explain the findings of increased pulmonary responsiveness to hypoxia, or the tendency to

venous thromboses seen in Chuvash patients. I did not see a difference in levels of *EPO* expression in the mutant cells, in keeping with our findings with the other genes. This may be due to low levels of *HIF-2A* and *EPO* expression seen in lymphoblastoid cells [336], however it stands in contrast to the strikingly high levels of EPO seen in the patient himself [287]. It therefore highlights the importance of tissue-specific EPO expression.

Previous studies have had variable success showing up-regulation of HIF-target genes, with some studies showing significant normoxic induction in all genes tested and others unable to do so [322]. As mentioned previously, there are problems with using EBV-transformed B lymphocytes, particularly the unquantifiable effects of transformation and prolonged culture. However this system is relatively easy to perform, with minimal inconvenience to the patient and control donors. An additional caveat is that the existence of increased levels of HIF-1α protein does not necessarily mean that it will be transcriptionally active.

Contrary to expectations, the gene expression data suggest that these mutations in *VHL* affect neither HIF-target gene expression at baseline, not its induction by hypoxia. These results show a high degree of internal consistency, and have sufficient power to support the null hypothesis that there is no significant difference between this case and controls.

6.3.04 HIF regulation

The central question remains, therefore, why we do not see stabilisation of HIF-1α in patient-derived LCL cells, given that expression of VHL p19 is reduced in these cells and HIF-1a does appear to be stabilised in RCC-10 cells transfected with the individual mutants? The pattern of ubiquitination seen in mutant LCL cells appears to be abnormal, suggesting that VHL's E3 ligase activity is in some way compromised. One possibility is that in these LCL cells there is another mechanism at work regulating HIF-1α levels. However, HIF-1α was stabilised by inhibition of the PHD enzymes by DFO and by the use of proteasomal inhibitors, implying that this alternative mechanism requires both prolyl hydroxylation and a functional proteasome.

Previous studies have shown stabilisation of HIF-1α in patient–derived LCL cells [317, 333], although not all studies of hereditary erythrocytosis have been able to use LCL cells to show activation of the oxygen sensing pathway [322]. This case presented with a much more severe clinical phenotype than has been observed in other cases of hereditary erythrocytosis. Patients with Chuvash polycythaemia do exhibit a reduced life expectancy, but live well into

their 4th decade [319]. Moreover, the small number of patients studied so far with mutations in *PHD2* or *HIF2A* have also presented in adulthood [127, 322, 328]. We might therefore expect that the D126N and S183L mutants might produce greater activation of the oxygen sensing pathway.

I was also unable to show activation of HIF-2α. HIF-2α is not highly expressed in peripheral blood mononuclear cells or LCL cells [336]. Given the evidence outlined in the introduction to this chapter that HIF-2α has a critical role in EPO control in humans, this is a significant limitation in the usefulness of LCL cells for the study of hereditary erythrocytosis.

The ubiquitin-proteasome system is the major cytosolic proteolytic system in eukaryotes; the specificity of the system is determined by the numerous E3 ligases that mediate substrate recognition [345]. Polyubiquitin chains linked through lysine 48 are the principal signal for targeting substrates to the 26S proteasome. Longer ubiquitin chains seem to associate more tightly with the proteasome, but the minimal signal for efficient proteasomal targeting is a tetra-ubiquitin [346]. However, protein modification by ubiquitin is involved in the regulation of numerous cellular functions besides protein stability. Ubiquitin is a highly versatile molecular signal because of its ability to modify substrate proteins in its monomeric form or to be conjugated to preceding ubiquitin moieties in a variety of orientations to form many types of ubiquitin chain [347]. Following binding to VHL, HIF-1α is ubiquitinated on three conserved lysines within the oxygen dependent degradation domain [300]. It is then targeted for destruction by the 26S proteasome [301]. Very little is known about the type and length of ubiquitination that is required for this process. Moreover it is not known whether HIF-α ubiquitination can be associated with any function other than degradation.

These results show that the pattern of ubiquitination of HIF-1α in the mutant LCL cells is abnormal. They suggest that mutant cells are unable to form long-chain poly-ubiquitinated molecules due to an impairment of VHL's ubiquitin ligase activity. This deficit is not associated with stabilisation of HIF-1α in normoxia, however, or with up regulation of HIF target genes.

There exist a number of possibilities why HIF-1α is not stabilised in these cells. First, there is the technical issue of using just two transformations of LCL cells from this patient. An important positive control would be LCL cells carrying the Chuvash mutation, or indeed other mutations associated with hereditary erythrocytosis. It would also be interesting to study

heterozygous cells derived from the patient's parents, however if each mutation results in a mild loss of function, the remaining normal allele may entirely compensate for such deficiency.

Second, there may be a threshold effect, since HIF degradation is an extremely rapid process [125]. The length of ubiquitin chain may not affect proteasomal degradation significantly enough to result in functional consequences *in vitro*. In patients with classical VHL disease somatic inactivation of the remaining wild-type allele results in constitutive HIF activation. By contrast, erythrocytosis is hypothesised to arise from mild activation of the HIF pathway in every cell. Lentiviral knockdown of *VHL* resulted in normoxic stabilisation of HIF in LCL cells, with reduction of the polyubiquitinated HIF-1α species in control cells seen in hypoxia or following proteasomal inhibition. Moreover, despite these effects, knockdown of *VHL* did not lead to changes in HIF target gene expression. This implies that the D126N and S183L mutations may simply result in a loss of function below the threshold for detection.

An alternative possibility is that these results expose a novel mechanism of HIF-1α regulation. Although VHL results in rapid degradation of HIF-1α in normoxia, there is evidence of HIF-1α turnover in *VHL* -/- cells treated with hypoxia and DFO [348], in cells transfected with HIF-1α that has had both hydroxyl-acceptor prolines [349] or all three ubiquitin-acceptor lysines mutated [300], and in cells with an inactive E1 enzyme [348]. The receptor of activated protein kinase C 1 (RACK1) has been shown to compete with HSP90 for HIF-1α binding, and results in VHL and oxygen-independent proteasomal degradation [350]. Activation of p53 by DNA damage leads to p53 E3 ubiquitin protein ligase homolog (MDM2)-mediated HIF ubiquitination and destruction by the proteasome [351].

A number of proteins show degradation through both ubiquitin-dependent and ubiquitin-independent pathways, including p53 and p21 [352]. There is evidence of ubiquitin-independent HIF-1α degradation in VHL -/- cells with an inactive E1 enzyme [348]. COMMD1 (copper metabolism (Murr1) domain containing 1) has been shown to complex with HIF-1α and HSP70 (heat shock 70kDa protein), resulting in ubiquitin-independent degradation by the 20S proteasome [353]. By analogy with MDM2 it is possible that VHL may promote direct binding to the proteasome, and that this may require different regions than those required for ubiquitin ligase activity [354]. Certainly, direct association of other SCF complexes with the proteasome has been observed [345], and there is some work to suggest that HIF-1α is able to bind directly to the PSMA7 subunit of the 20S proteasome [355]. Intriguingly, this is the same

subunit that has been shown to interact with a number of other cellular and viral proteins, including p21 [356] and EBNA3 [357].

Finally, it may be that another function of VHL entirely is responsible for the phenotype observed in this patient. The type 2C VHL mutations that are associated with phaeochromocytoma are normal with respect to HIF regulation [358], and may instead cause disease through effects on JunB [359] or fibronectin [308]. Despite the clear relationship between HIF and EPO, it is less obvious how abnormal VHL function contributes to pulmonary hypertension. Whilst lymphoblastoid cells offer a highly convenient model system with which to study the effects of these mutations, it is possible that another cell type, such as dermal fibroblasts or even induced pluripotent stem cells derived from patient urothelial cells, could provide more information.

6.4 Conclusions and future work

In conclusion, I have characterised the effect of two novel *VHL* mutations on the oxygen sensing pathway. These mutations result in reduced VHL protein stability and impair HIF-1α degradation in both RCC-10 cells and patient-derived lymphoblastoid cells. In particular, these mutations result in profoundly abnormal ubiquitination of the HIF-1α protein, a phenotype which can be recapitulated by lentiviral mediated knockdown of VHL in lymphoblastoid cells. However, despite this abnormal ubiquitination, HIF-1α is not stabilised in normoxia and there is no evidence of activation of HIF-target genes. This may represent the activity of a novel, VHL-independent, HIF regulatory pathway.

There are two main routes to take this work forward. First, new technologies, such as zinc-finger nucleases or Transcription Activator-Like Effector Nucleases (TALENs), offer the potential to produce a complete *VHL* knockout in lymphoblastoid cells. This would allow us to determine whether VHL ubiquitin ligase activity exhibited a threshold effect. It would also enable the re-expression of individual mutants in these cells. A second goal would be to generate a mouse model of VHL depletion in the B cell lineage. There is evidence for the importance of VHL function in T cells, and it would be intriguing to see if B cells did indeed harbour a novel HIF regulatory pathway.

Chapter 7 Discussion

7.1 Thesis summary

This thesis describes four renal disorders that are caused by the inheritance of a genetic variant. These disorders have arisen in different populations and are caused by diverse biochemical mechanisms. In each case, renal disease is due to a rare mutation that may be unique to that family. Many of these rare mutations originate in a single founder and are dispersed by his descendants in the population. The population frequency of the mutation is therefore related to the age of the mutation [197]. Successful linkage analysis requires that a mutation be highly penetrant, in order that the phenotype and thus affection status, can be determined accurately [48]. This helps to understand the mechanistic basis of these conditions, since highly penetrant disease is likely to be caused by a significant disturbance of the relevant biochemical pathway.

In chapter 3 I have studied the occurrence of renal disease in the Turkish-Cypriot population of Northern Cyprus. This small population has a high incidence of renal disease, much of which is undiagnosed and may be familial in origin. Recent reports have shown evidence of significant geographical clustering for a number of monogenetic conditions in Greek-Cypriots, which is consistent with a degree of population stratification. Given the genetic similarities between Greek- and Turkish-Cypriot populations on the island, I hypothesised that we would find evidence of the same renal conditions in the Turkish-Cypriot ESRD population that are found in the Greek-Cypriot population.

I identified three individuals with the G871C mutation in *COL4A3*, but no cases of *CFHR5* nephropathy, the G1334E mutation in *COL4A4,* or the 3854delG mutation in *COL4A4*. Two individuals with the G871C mutation were part of a large pedigree with renal impairment and microscopic haematuria, but the third was not known to be related. I then used conflicting homozygosity analysis to demonstrate a minimal haplotype that was shared between four Cypriot families with this mutation, thus dating this mutation to 17 generations ago. This is in agreement with evolutionary theory which argues that pathogenic variants with large effect size are of recent origin.

In chapter 4 I sought to identify a novel genetic cause of renal disease a large Greek-Cypriot kindred. This family had a very high incidence of renal disease, consistent with autosomal dominant inheritance, but with a variable clinical phenotype on renal biopsy. I used linkage analysis to identify a 15 cM region on chromosome 10, with a maximal LOD score of 4.187,

that contained no genes currently associated with renal disease. Next generation sequencing was performed in two first cousins from this pedigree, which identified six novel, non-synonymous coding variants within this region. Although these variants were described as novel by database comparison, I found that three of them occurred with significant frequency in our reference population of Turkish-Cypriots with and without renal disease.

I assessed the expression of three variants identified by next generation sequencing in primary cells derived from individuals in this pedigree. I studied renal tubular urothelial cells. These cells are easy to obtain, enable non-invasive study of renal tissue, and have the potential to generate induced pluripotent stem cells. This chapter demonstrates the use of discrete filtering steps to identify private variants in this family that may be responsible for renal disease. The next step is to show by what mechanism one of these variants is responsible for renal disease.

In chapter 5 I studied another large pedigree with adult-onset renal failure and a variable clinical phenotype on renal biopsy. In this instance, linkage analysis was unable to demonstrate co-segregation of a single genomic region with renal disease. This could have been due to polygenic inheritance, the presence of phenocopies within the pedigree, or an alternative mode of inheritance. Sequencing of the mitochondrial genome in this family demonstrated the presence of a novel polymorphism in the HSP1 region at homoplastic levels. Subsequent examination of mitochondrial function in primary fibroblasts demonstrated a significant reduction in baseline oxidative respiration with a compensatory increase in glycolysis. Mitochondrial disease is a source of non-chromosomal inheritance, and has recently been implicated in a range of complex diseases, including renal disease. As such, mitochondrial disorders may be an important source of the 'missing heritability' that renal GWAS have failed to identify.

Lastly, in chapter 6 I have taken a distinctive phenotype in a single individual and related it to perturbations in the well-studied oxygen sensing pathway mediated by *VHL* and *HIF*. This phenotype was so distinctive that it was possible to use candidate gene sequencing to identify the causative variant [287]. However, *VHL* mutations exhibit a high degree of allelic heterogeneity which may reflect the ability of VHL to interact with and regulate a wide range of biological partners. Recent studies suggest that erythrocytosis associated with *VHL* mutations is due to the combination of two hypomorphic alleles, resulting in partial activation

of HIF [286]. However the variants that I studied in patient-derived B-cell lines showed abnormal degradation of HIF without activation of HIF target genes. It is therefore possible that these cells are able to employ some kind of VHL-independent HIF regulatory mechanism.

All these studies demonstrate, in differing ways, the challenges of linking phenotype to genotype. It is difficult to study renal pathology directly, and so we are forced to use a series of surrogate markers with variable relevance to the underlying physiology. Where the phenotype is not distinctive, this increases the experimental 'noise', and reduces the strength of genetic association or linkage. Alternatively, the lack of association can indicate the contribution of multiple genes, or the existence of non-chromosomal inheritance. The hope, of course, is that new genetic discoveries will uncover novel genetic, physiological, and therapeutic pathways. The incredible power of next generation sequencing technologies now makes it possible to determine all the possible causative variants within a few weeks. However, identification of the underlying biochemical mechanisms is likely to remain a considerable challenge for the future.

7.2 The GIKD study

Many of the DNA samples studied in this thesis were collected as part of the genetic investigation of kidney disease (GIKD) study (UKCRN Portfolio ID: 3946). This study was set up in 2007 in order to identify the genetic and molecular basis of inherited kidney disease. The inclusion criteria were simply the presence of renal disease that appeared to be transmitted genetically, but for which there was no established genetic diagnosis. A secondary objective was the creation of a bank of DNA and blood samples to allow future investigations into these or related diseases.

In its first inception the GIKD study was designed to recruit large pedigrees that could be solved by linkage analysis. An important part of the recruitment process was the identification of affected relatives who might not be known to the renal team at the recruiting centre. This process was facilitated by the recommendation in 2008 from the national institute of clinical excellence (NICE; Clinical guidelines, CG73) that testing for CKD should be offered to all those with a family history of stage 5 CKD. The expectation was that with adequately sized pedigrees

it would be possible to use SNP linkage studies to identify a region of the genome in which to look for a causative mutation.

The identification of *CFHR5* nephropathy was the first success for the GIKD study [137]. The findings of Gale *et al.* illustrate a number of issues that have been important to the studies in this thesis. Most importantly, *CFHR5* nephropathy exhibits a highly unusual renal phenotype, manifest as deposition of complement C3 without detectable immunoglobulin on renal biopsy. This made the existence of phenocopies highly unlikely, and therefore it was possible to combine two pedigrees together with reasonable confidence. Further confidence to combine pedigrees came from the close geographical origin of these two pedigrees, suggesting that a single mutation had arisen in a common ancestor. By combining pedigrees it was possible to generate a significant LOD score and exclude a large portion of the genome from further analysis.

Prior to the widespread use of next generation sequencing, the identification of a linked interval left a large number of genes to be sequenced manually on a candidate by candidate basis. However, the renal phenotype in *CFHR5* nephropathy strongly suggested that the underlying abnormality was a gene directly involved in complement regulation. Consequently the best candidate genes for further study were associated with the regulator of complement activation gene cluster on chromosome 1. Finally, this disorder arose from an internal duplication, rather than a point mutation, and as such exhibited a normal Sanger sequencing sequence.

To date, the GIKD study has recruited samples from 145 pedigrees across four hospital sites in England. Dr Gale and I have performed SNP genotyping on 15 of the largest pedigrees, including the three discussed in this thesis. A significant limitation of the study has been the question of how to define the renal phenotype. Renal biopsy provides the best phenotypic marker in the absence of extra-renal signs, but may not have been performed even in those with ESRD, and may not include all diagnostic techniques, such as electron microscopy. In the absence of a clear renal phenotype linkage analysis can easily be confounded by phenocopies, as discussed in chapter 3. Finally, even with significant LOD scores, there remains the question of candidate gene sequencing, which can be expensive, time consuming, and may even miss the causative variant.

Fortunately, next generation sequencing has now become a standard technique for this kind of analysis. We have had over 40 exomes sequenced, with discovery of the likely pathogenic mutation in 14. Although the identification of novel pathogenic mutations requires exome data from more than one individual per pedigree (in the absence of other informative data), it is possible to uncover known causative mutations with the sequence data from just one individual. In this way it is possible to screen pedigrees with a limited renal phenotype for mutations in genes associated with, for example, microscopic haematuria or focal segmental glomerulosclerosis. Given the expense of commercial sequencing for individual genes (£670 each for *COL4A3* and *COL4A4*), it makes economic sense to pursue next generation sequencing in such cases (£800 per exome). Moreover, once the index case has been sequenced it is then possible to decide whether it is appropriate to sequence additional individuals.

With such a tsumami of information [360], the question turns to the more difficult one of proving that a given variant is pathogenic. There have been a few ethical amendments to the GIKD study protocol that may assist with this. First, the need to screen relatives was addressed early on, with the design of a template that could be sent out to relatives for them to take to their local GP. Second, it is important to confirm that all affected individuals in a pedigree carry the same variant. Occasionally this can only be ascertained by means of examining archived histological tissue, or even by testing stored DNA samples (for example if that individual was on the transplant register) after the individual had died. Finally, it becomes increasingly important to study the effect of variants *in vitro*. Dermal fibroblasts are commonly used for this purpose, as discussed in chapter 5, but it is preferable to look at the function of renal cells. Urothelial cell culture is a step in this direction (chapter 4), and the ability to transform these cells into pluripotent stem cells offers even greater promise [161].

7.3 Personalised genomic medicine

The studies illustrated in this thesis provide an example of the direction that personalised genomic medicine is moving in. Within a decade, the scientific community has moved from the collaborative effort and cost that resulted in the first complete human genomic sequence to the generation of equivalent volumes of data by one lab within weeks for just a few hundred dollars [70]. Given the importance of the information thus generated it is inevitable that this process will move into the clinical sphere, as suggested by Sir John Bell's recent report

to the Human Genomics Strategy Group [361]. The question, then, is how can this information best be used and what ethical issues need to be addressed for the optimal use of this data.

Gene discovery is the starting point both for understanding the genetic mechanisms underlying diseases and for providing clues to therapeutic approaches. In the short term, the ability to classify diseases into sub-phenotypes from genetic information may result in improved treatment and an expanded use of pharmacogenomics. This is especially true for cancer, where combinations of drugs could be used to target the abnormal molecular pathways seen within each tumour. Drug treatments might be anticipated to show larger treatment effects where treatment is matched to the underlying genetic cause. Thus the ability to stratify individuals according to genotype has the potential to make clinical trials more cost-effective and time-efficient by enrolling a much smaller number of patients [59]. Even now, the ability to differentiate between genetic and non-genetic causes of the nephrotic syndrome in children has the possibility to influence treatment options.

In the longer term, the discovery of further genetic risk factors for kidney disease will facilitate the discovery of new drug targets through the understanding of novel molecular pathways. The use of selective inhibition of VEGF and other angiogenic growth factors in *VHL*-driven kidney cancer is exactly the kind of rational drug discovery that is possible with a detailed understanding of the genetic defect in a given renal disease [19]. The major uncertainty here is the extent to which rare and unique mutations mediate their effects through common genetic pathways that are amenable to drug development. To put this in context, the US Food and Drug Administration define orphan diseases as those occurring in less than 200,000 cases in the United States, and have adopted this definition because it is roughly the cut-off point for willingness by pharmaceutical companies to work on treatments for a disease [362]. Significant changes will be required to the partnership between academia and industry in this regard to consider targeted drug developments for very rare diseases [362].

Another long-term consequence of personalised genomic medicine is the ability to predict disease susceptibility. There are two areas in which this may come about. First, there is the question of finding the genetic modifiers that influence a person's risk of disease [57]. Oligogenic inheritance has been demonstrated *in vitro* for the ciliopathies [87], and work is on-going to uncover factors that influence the risk of disease over and above the simple additive model looked at by GWAS studies. Second, as our understanding of genetic disease grows it

will become increasingly possible to study a healthy person's genome for evidence of rare mutations that confer a significant risk disease. Given that we each contain around 175 *de novo* mutations [78], it is likely that this kind of predictive information is relevant to us all.

However, the ability to generate and share large amounts of individual-specific information by next-generation sequencing generates new manifestations of perennial ethical issues such as consent, feedback, protection of privacy and the governance of research [363-365]. Two areas of research ethics are especially pertinent to the current discussion: the limitations of the current consent process and the management of individual research results.

Whilst informed consent has been a guiding principle of medical investigation since the mid-20th century, genomics research moves away from this principle in several respects [71]. First, DNA provides information not just on an individual, but also on their relatives and ethnic group in a format that is easily accessible across international borders. Second, the quantity of information available from DNA is such that even with personal identifiers removed it may be possible to identify the original donor. Third, DNA is often archived in online repositories, such as mutation and variation databases. These are increasingly important for the wider scientific community; however the complexity of genomics research along with the difficulty of providing precise specifications for the future use of data raise questions as to whether consent can be truly 'informed'.

At an individual level, an important ethical issue is the question of reporting findings back to the participant [71]. Most study participants are motivated by trying to discover what is responsible for the disease in their family, whether they themselves are affected or not. However the question arises how to manage findings that are incidental. There appears to be some agreement that, where there is a serious treatable condition, researchers have a moral obligation to feed this information back to research participants [366]. However, where findings are of a less serious nature, are untreatable or of uncertain significance, the potential benefits for participants of being informed need to be balanced against the participant's right not to know.

Final issues for personalised genomic medicine are the questions of cost and of data storage. The cost of genome sequencing has fallen by several orders of magnitude over the last decade. It has now reached the point where it will soon undercut the current cost of NHS accredited genetic testing. If whole genome sequencing is to replace current genetic testing, not only will

staff need the technical and ethical training to ensure proper conduct and competence to practice, but clinical staff will need substantial bioinformatics support. Whilst the cost of data storage has also fallen dramatically, there remain issues of how to back the data up and how to ensure that it is accessible for analysis and collaboration while safeguarding aspects of patient privacy. Automated data management and storage provides the most efficient use of resources, but has additional implications for data security and confidentiality. Given the density of information storage on the DNA molecule, this may even form the most durable and cost-effective long-term information store [367] .

7.4 Future directions

Current sequencing technologies have revolutionised the field of medical genetics. It is increasingly affordable to generate volumes of data that are many orders of magnitude greater and less expensive than they were just a decade ago. Whilst technical advances have revolutionised the process of sequence acquisition, there remain several areas where focused efforts are likely to advance the field substantially [73].

The first is the more accurate description of phenotypes, particularly in the context of Mendelian disorders. As has been discussed repeatedly in this thesis, the definition of accurate phenotypes is of central importance in all kinds of genetic analysis. There are hundreds, perhaps thousands of poorly defined familial phenotypes that are rare or unique, especially in high risk or non-European populations [73]. The discovery of the underlying genes would be greatly facilitated by the detailed clinical investigation of such cases. Moreover, collaborative storage of such descriptive information in international repositories would facilitate the delineation of new Mendelian disorders and allow samples from similar phenotypes to be combined.

For non-Mendelian conditions, the challenge is to define phenotypes that would be amenable to study. Examples of suitable renal phenotypes include acute kidney injury or transplant rejection [47]. A refinement of this technique is the search for genetic modifiers, such as those that influence the speed of progression of chronic kidney disease. These might be studied in the context of a defined genetic background, such as ADPKD or mutations in *COL4A3/A4*.

Another issue is the need for improved technical, statistical and bioinformatic methods for data analysis. Technical issues related to current next generation sequencing protocols include the methods for reducing false-positive and false-negative variant calls, calling indels, prioritising candidate causal variants, and predicting and annotating the potential functional impact for disease gene discovery or molecular diagnostics [73]. Statistical issues include the need for models of non-additive genetic effects, including gene-gene and gene-environmental interaction. Such technical issues may pose a hurdle for clinical laboratory accreditation [71]. That said however, next generation sequencing is likely to be introduced into the clinical setting before these challenges are fully resolved owing in part to their ability to facilitate diagnosis and inform therapy [360, 368, 369].

It is now a realistic goal to identify a candidate gene for every recessive disorder using as few as one affected individual per disorder [73]. Solving all dominant disorders will be more difficult, requiring technical and analytic improvements that should take place over the next few years. This will require the coordination and cooperation of the human genetics community and funding agencies. However, efforts are already underway at multiple centres throughout the world to establish the collaborative framework and physical infrastructure required. Even after the genetic basis of all Mendelian disorders has been identified [73], work will still be required to understand disease mechanisms better and develop improved therapeutics. Given the rate of *de novo* mutations, however, the diagnostic challenge may never entirely fade away.

Human genetic disorders therefore provide us with an extraordinarily large array of natural experiments. Given the population of the planet, it is probable that there exist individuals with mutations in the vast majority of coding genes. Technology now affords us the chance to identify markers and even the individual variant that is associated with every phenotype that we encounter. With the techniques of association so well established, research is now required to demonstrate how a given variant is sufficient to cause a disease. This thesis traces the path of gene identification from its origins in a human population, and offers a range of insights into how causality can be demonstrated. This step is now the most difficult, and therefore has become limiting in the analysis of the genetic basis of kidney disease.

Chapter 8 References

References

1. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.

2. Dunham, I., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

3. Jones, C.A., et al., *Serum creatinine levels in the US population: third National Health and Nutrition Examination Survey.* Am J Kidney Dis, 1998. **32**(6): p. 992-9.

4. Quinn, M.P., et al., *The practical implications of using standardized estimation equations in calculating the prevalence of chronic kidney disease.* Nephrol Dial Transplant, 2008. **23**(2): p. 542-8.

5. Hallan, S.I., et al., *International comparison of the relationship of chronic kidney disease prevalence and ESRD risk.* J Am Soc Nephrol, 2006. **17**(8): p. 2275-84.

6. Levey, A.S., et al., *Chronic kidney disease: common, harmful, and treatable--World Kidney Day 2007.* J Am Soc Nephrol, 2007. **18**(2): p. 374-8.

7. Go, A.S., et al., *Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization.* N Engl J Med, 2004. **351**(13): p. 1296-305.

8. Johnson, E.S., et al., *Predicting renal replacement therapy and mortality in CKD.* Am J Kidney Dis, 2007. **50**(4): p. 559-65.

9. Caskey, F.J., et al., *Global variation in renal replacement therapy for end-stage renal disease.* Nephrol Dial Transplant, 2011. **26**(8): p. 2604-10.

10. Bash, L.D., B.C. Astor, and J. Coresh, *Risk of incident ESRD: a comprehensive look at cardiovascular risk factors and 17 years of follow-up in the Atherosclerosis Risk in Communities (ARIC) Study.* Am J Kidney Dis, 2010. **55**(1): p. 31-41.

11. Hallan, S.I., et al., *Combining GFR and albuminuria to classify CKD improves prediction of ESRD.* J Am Soc Nephrol, 2009. **20**(5): p. 1069-77.

12. De Vecchi, A.F., M. Dratwa, and M.E. Wiedemann, *Healthcare systems and end-stage renal disease (ESRD) therapies--an international review: costs and reimbursement/funding of ESRD therapies.* Nephrol Dial Transplant, 1999. **14 Suppl 6**: p. 31-41.

13. Boddana, P., et al., *UK Renal Registry 11th Annual Report (December 2008): Chapter 14 UK Renal Registry and international comparisons.* Nephron Clin Pract, 2009. **111 Suppl 1**: p. c269-76.

14. USRDS. *U.S. Renal Data System, USRDS 2011 Annual Data Report: Atlas of End-Stage Renal Disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2011.* 2008 2012]; Available from: http://www.usrds.org/adr_2008.htm.

15. ERA. *ERA-EDTA Registry Annual Report 2008*. 2008 [cited 2011; Available from: http://www.era-edta-reg.org/index.jsp?p=1.

16. Courtney, A.E., P.T. McNamee, and A.P. Maxwell, *The evolution of renal transplantation in clinical practice: for better, for worse?* QJM, 2008. **101**(12): p. 967-78.

17. Freedman, B.I., et al., *Population-based screening for family history of end-stage renal disease among incident dialysis patients.* Am J Nephrol, 2005. **25**(6): p. 529-35.

18. UK, C.R., *Kidney Cancer - UK Incidence Statistics*, 2011.

19. Linehan, W.M., et al., *Molecular diagnosis and therapy of kidney cancer.* Annu Rev Med, 2010. **61**: p. 329-43.

20. Maher, E.R., *Genetics of familial renal cancers.* Nephron Exp Nephrol, 2011. **118**(1): p. e21-6.

21. Ricketts, C., et al., *Germline SDHB mutations and familial renal cell carcinoma.* J Natl Cancer Inst, 2008. **100**(17): p. 1260-2.

22. Doris, P.A., *Genetic susceptibility to hypertensive renal disease.* Cell Mol Life Sci, 2012. **69**(22): p. 3751-63.

23. Hildebrandt, F., *Genetic kidney diseases.* Lancet, 2010. **375**(9722): p. 1287-95.

24. McKnight, A.J., D. Currie, and A.P. Maxwell, *Unravelling the genetic basis of renal diseases; from single gene to multifactorial disorders.* J Pathol, 2010. **220**(2): p. 198-216.

25. Benoit, G., et al., *Hereditary kidney diseases: highlighting the importance of classical Mendelian phenotypes.* Ann N Y Acad Sci, 2010. **1214**: p. 83-98.

26. Wang, Y.Y. and J. Savige, *The epidemiology of thin basement membrane nephropathy.* Semin Nephrol, 2005. **25**(3): p. 136-9.

27. Harris, P.C. and V.E. Torres, *Polycystic kidney disease.* Annu Rev Med, 2009. **60**: p. 321-37.

28. Arts, H.H. and N.V. Knoers, *Current insights into renal ciliopathies: what can genetics teach us?* Pediatr Nephrol, 2012.

29. Rao, M., I. Peter, and T.A. Trikalinos, *A lesson from the Zuni Indians: heritability in perspective.* Am J Kidney Dis, 2010. **56**(2): p. 251-4.

30. Visscher, P.M., W.G. Hill, and N.R. Wray, *Heritability in the genomics era--concepts and misconceptions.* Nat Rev Genet, 2008. **9**(4): p. 255-66.

31. Fox, C.S., et al., *Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study.* J Am Soc Nephrol, 2004. **15**(9): p. 2457-61.

32. Bochud, M., et al., *Heritability of renal function in hypertensive families of African descent in the Seychelles (Indian Ocean).* Kidney Int, 2005. **67**(1): p. 61-9.

33. Forsblom, C.M., et al., *Heritability of albumin excretion rate in families of patients with Type II diabetes.* Diabetologia, 1999. **42**(11): p. 1359-66.

34. Fogarty, D.G., et al., *Urinary albumin excretion in families with type 2 diabetes is heritable and genetically correlated to blood pressure.* Kidney Int, 2000. **57**(1): p. 250-7.

35. Langefeld, C.D., et al., *Heritability of GFR and albuminuria in Caucasians with type 2 diabetes mellitus.* Am J Kidney Dis, 2004. **43**(5): p. 796-800.

36. Placha, G., et al., *A genome-wide linkage scan for genes controlling variation in renal function estimated by serum cystatin C levels in extended families with type 2 diabetes.* Diabetes, 2006. **55**(12): p. 3358-65.

37. MacCluer, J.W., et al., *Heritability of measures of kidney disease among Zuni Indians: the Zuni Kidney Project.* Am J Kidney Dis, 2010. **56**(2): p. 289-302.

38. Traglia, M., et al., *Heritability and demographic analyses in the large isolated population of Val Borbera suggest advantages in mapping complex traits genes.* PLoS One, 2009. **4**(10): p. e7554.

39. Rudan, I., et al., *Inbreeding and risk of late onset complex disease.* J Med Genet, 2003. **40**(12): p. 925-32.

40. Bittles, A.H. and J.V. Neel, *The costs of human inbreeding and their implications for variations at the DNA level.* Nat Genet, 1994. **8**(2): p. 117-21.

41. Charlesworth, B. and D. Charlesworth, *The genetic basis of inbreeding depression.* Genet Res, 1999. **74**(3): p. 329-40.

42. Lei, H.H., et al., *Familial aggregation of renal disease in a population-based case-control study.* J Am Soc Nephrol, 1998. **9**(7): p. 1270-6.

43. Satko, S.G. and B.I. Freedman, *The importance of family history on the development of renal disease.* Curr Opin Nephrol Hypertens, 2004. **13**(3): p. 337-41.

44. Seaquist, E.R., et al., *Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy.* N Engl J Med, 1989. **320**(18): p. 1161-5.

45. Freedman, B.I., et al., *Genetic factors in diabetic nephropathy.* Clin J Am Soc Nephrol, 2007. **2**(6): p. 1306-16.

46. Fioretto, P., et al., *Is diabetic nephropathy inherited? Studies of glomerular structure in type 1 diabetic sibling pairs.* Diabetes, 1999. **48**(4): p. 865-9.

47. Kottgen, A., *Genome-wide association studies in nephrology research.* Am J Kidney Dis, 2010. **56**(4): p. 743-58.

48. Dawn Teare, M. and J.H. Barrett, *Genetic linkage studies.* Lancet, 2005. **366**(9490): p. 1036-44.

49. Burton, P.R., M.D. Tobin, and J.L. Hopper, *Key concepts in genetic epidemiology.* Lancet, 2005. **366**(9489): p. 941-51.

50. Nussbaum, R.L., et al., *Thompson & Thompson genetics in medicine*. 7th ed2007, Philadelphia: Saunders/Elsevier. xi, 585 p.

51. Morton, N.E., *Sequential tests for the detection of linkage.* Am J Hum Genet, 1955. **7**(3): p. 277-318.

52. Chotai, J., *On the lod score method in linkage analysis.* Ann Hum Genet, 1984. **48**(Pt 4): p. 359-78.

53. Lander, E. and L. Kruglyak, *Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results.* Nat Genet, 1995. **11**(3): p. 241-7.

54. Hall, J.M., et al., *Linkage of early-onset familial breast cancer to chromosome 17q21.* Science, 1990. **250**(4988): p. 1684-9.

55. Wooster, R., et al., *Identification of the breast cancer susceptibility gene BRCA2.* Nature, 1995. **378**(6559): p. 789-92.

56. Duggirala, R., et al., *A variance component approach to dichotomous trait linkage analysis using a threshold model.* Genet Epidemiol, 1997. **14**(6): p. 987-92.

57. Bodmer, W. and C. Bonilla, *Common and rare variants in multifactorial susceptibility to common diseases.* Nat Genet, 2008. **40**(6): p. 695-701.

58. O'Seaghdha, C.M. and C.S. Fox, *Genome-wide association studies of chronic kidney disease: what have we learned?* Nat Rev Nephrol, 2012. **8**(2): p. 89-99.

59. Frazer, K.A., et al., *Human genetic variation and its contribution to complex traits.* Nat Rev Genet, 2009. **10**(4): p. 241-51.

60. Risch, N. and K. Merikangas, *The future of genetic studies of complex human diseases.* Science, 1996. **273**(5281): p. 1516-7.

61. Hattersley, A.T. and M.I. McCarthy, *What makes a good genetic association study?* Lancet, 2005. **366**(9493): p. 1315-23.

62. Kao, W.H., et al., *MYH9 is associated with nondiabetic end-stage renal disease in African Americans.* Nat Genet, 2008. **40**(10): p. 1185-92.

63. Kopp, J.B., et al., *MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis.* Nat Genet, 2008. **40**(10): p. 1175-84.

64. Genovese, G., et al., *Association of trypanolytic ApoL1 variants with kidney disease in African Americans.* Science, 2010. **329**(5993): p. 841-5.

65. Klein, R.J., et al., *Complement factor H polymorphism in age-related macular degeneration.* Science, 2005. **308**(5720): p. 385-9.

66. Stanescu, H.C., et al., *Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy.* N Engl J Med, 2011. **364**(7): p. 616-26.

67. Kottgen, A., et al., *New loci associated with kidney function and chronic kidney disease.* Nat Genet, 2010. **42**(5): p. 376-84.

68. Levy, S., et al., *The diploid genome sequence of an individual human.* PLoS Biol, 2007. **5**(10): p. e254.

69. Wheeler, D.A., et al., *The complete genome of an individual by massively parallel DNA sequencing.* Nature, 2008. **452**(7189): p. 872-6.

70. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.

71. Majewski, J., et al., *What can exome sequencing do for you?* J Med Genet, 2011. **48**(9): p. 580-9.

72. Gilissen, C., et al., *Disease gene identification strategies for exome sequencing.* Eur J Hum Genet, 2012. **20**(5): p. 490-7.

73. Bamshad, M.J., et al., *Exome sequencing as a tool for Mendelian disease gene discovery.* Nat Rev Genet, 2011. **12**(11): p. 745-55.

74. Ng, S.B., et al., *Targeted capture and massively parallel sequencing of 12 human exomes.* Nature, 2009. **461**(7261): p. 272-6.

75. Montenegro, G., et al., *Exome sequencing allows for rapid gene identification in a Charcot-Marie-Tooth family.* Ann Neurol, 2011. **69**(3): p. 464-70.

76. Tong, M.Y., C.A. Cassa, and I.S. Kohane, *Automated validation of genetic variants from large databases: ensuring that variant references refer to the same genomic locations.* Bioinformatics, 2011. **27**(6): p. 891-3.

77. Byrne, C., et al., *UK Renal Registry 12th Annual Report (December 2009): chapter 3: UK ESRD incident rates in 2008: national and centre-specific analyses.* Nephron Clin Pract, 2010. **115 Suppl 1**: p. c9-39.

78. McClellan, J. and M.C. King, *Genetic heterogeneity in human disease.* Cell, 2010. **141**(2): p. 210-7.

79. Nachman, M.W. and S.L. Crowell, *Estimate of the mutation rate per nucleotide in humans.* Genetics, 2000. **156**(1): p. 297-304.

80. *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, 1994. **77**(6): p. 881-94.

81. Mochizuki, T., et al., *PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein.* Science, 1996. **272**(5266): p. 1339-42.

82. Harris, P.C., *2008 Homer W. Smith Award: insights into the pathogenesis of polycystic kidney disease from gene discovery.* J Am Soc Nephrol, 2009. **20**(6): p. 1188-98.

83. Rossetti, S., et al., *The position of the polycystic kidney disease 1 (PKD1) gene mutation correlates with the severity of renal disease.* J Am Soc Nephrol, 2002. **13**(5): p. 1230-7.

84. Magistroni, R., et al., *Genotype-renal function correlation in type 2 autosomal dominant polycystic kidney disease.* J Am Soc Nephrol, 2003. **14**(5): p. 1164-74.

85. Beales, P.L., et al., *Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome.* Am J Hum Genet, 2003. **72**(5): p. 1187-99.

86. Badano, J.L., et al., *Dissection of epistasis in oligogenic Bardet-Biedl syndrome.* Nature, 2006. **439**(7074): p. 326-30.

87. Zaghloul, N.A. and N. Katsanis, *Functional modules, mutational load and human genetic disease.* Trends Genet, 2010. **26**(4): p. 168-76.

88. Dedoussis, G.V., et al., *Co-inheritance of a PKD1 mutation and homozygous PKD2 variant: a potential modifier in autosomal dominant polycystic kidney disease.* Eur J Clin Invest, 2008. **38**(3): p. 180-90.

89. Vujic, M., et al., *Incompletely penetrant PKD1 alleles mimic the renal manifestations of ARPKD.* J Am Soc Nephrol, 2010. **21**(7): p. 1097-102.

90. Losekoot, M., et al., *Neonatal onset autosomal dominant polycystic kidney disease (ADPKD) in a patient homozygous for a PKD2 missense mutation due to uniparental disomy.* J Med Genet, 2012. **49**(1): p. 37-40.

91. Bergmann, C., et al., *Mutations in multiple PKD genes may explain early and severe polycystic kidney disease.* J Am Soc Nephrol, 2011. **22**(11): p. 2047-56.

92. Lantinga-van Leeuwen, I.S., et al., *Kidney-specific inactivation of the Pkd1 gene induces rapid cyst formation in developing kidneys and a slow onset of disease in adult mice.* Hum Mol Genet, 2007. **16**(24): p. 3188-96.

93. Piontek, K., et al., *A critical developmental switch defines the kinetics of kidney cyst formation after loss of Pkd1.* Nat Med, 2007. **13**(12): p. 1490-5.

94. Pei, Y., et al., *Somatic PKD2 mutations in individual kidney and liver cysts support a "two-hit" model of cystogenesis in type 2 autosomal dominant polycystic kidney disease.* J Am Soc Nephrol, 1999. **10**(7): p. 1524-9.

95. Qian, F., et al., *The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I.* Cell, 1996. **87**(6): p. 979-87.

96. Harris, P.C. and S. Rossetti, *Molecular diagnostics for autosomal dominant polycystic kidney disease.* Nat Rev Nephrol, 2010. **6**(4): p. 197-206.

97. King, K., et al., *Unusual deep intronic mutations in the COL4A5 gene cause X linked Alport syndrome.* Hum Genet, 2002. **111**(6): p. 548-54.

98. Rossetti, S., et al., *Incompletely penetrant PKD1 alleles suggest a role for gene dosage in cyst initiation in polycystic kidney disease.* Kidney Int, 2009. **75**(8): p. 848-55.

99. Rossetti, S., et al., *Identification of gene mutations in autosomal dominant polycystic kidney disease through targeted resequencing.* J Am Soc Nephrol, 2012. **23**(5): p. 915-33.

100. Maher, E.R., et al., *Von Hippel-Lindau disease: a genetic study.* J Med Genet, 1991. **28**(7): p. 443-447.

101. Maher, E.R., *Von Hippel-Lindau disease.* Curr Mol Med, 2004. **4**(8): p. 833-842.

102. Gruber, M.B., et al., *Papillary cystadenoma of epididymis: component of von Hippel-Lindau syndrome.* Urology, 1980. **16**(3): p. 305-306.

103. Neumann, H.P., et al., *Pancreatic lesions in the von Hippel-Lindau syndrome.* Gastroenterology, 1991. **101**(2): p. 465-471.

104. Kaelin, W.G., *Von Hippel-Lindau disease.* Annu Rev Pathol, 2007. **2**: p. 145-173.

105. Maher, E.R., et al., *Clinical features and natural history of von Hippel-Lindau disease.* Q J Med, 1990. **77**(283): p. 1151-1163.

106. Hough, D.M., et al., *Pancreatic lesions in von Hippel-Lindau disease: prevalence, clinical significance, and CT findings.* AJR Am J Roentgenol, 1994. **162**(5): p. 1091-1094.

107. Lubensky, I.A., et al., *Multiple neuroendocrine tumors of the pancreas in von Hippel-Lindau disease patients: histopathological and molecular genetic analysis.* Am J Pathol, 1998. **153**(1): p. 223-231.

108. Melmon, K.L. and S.W. Rosen, *Lindau's Disease. Review of the literature and study of a large kindred.* Am J Med, 1964. **36**: p. 595-617.

109. Lonser, R.R., et al., *von Hippel-Lindau disease.* Lancet, 2003. **361**(9374): p. 2059-2067.

110. Hes, F.J., C.J. Lips, and R.B. van der Luijt, *Molecular genetic aspects of Von Hippel-Lindau (VHL) disease and criteria for DNA analysis in subjects at risk.* Neth J Med, 2001. **59**(5): p. 235-243.

111. Klein, B., G. Weirich, and H. Brauch, *DHPLC-based germline mutation screening in the analysis of the VHL tumor suppressor gene: usefulness and limitations.* Hum Genet, 2001. **108**(5): p. 376-384.

112. Hes, F.J., et al., *Frequency of Von Hippel-Lindau germline mutations in classic and non-classic Von Hippel-Lindau disease identified by DNA sequencing, Southern blot analysis and multiplex ligation-dependent probe amplification.* Clin Genet, 2007. **72**(2): p. 122-129.

113. Neumann, H.P., et al., *Hemangioblastomas of the central nervous system. A 10-year study with special reference to von Hippel-Lindau syndrome.* J Neurosurg, 1989. **70**(1): p. 24-30.

114. Fukino, K., et al., *A family with hydrocephalus as a complication of cerebellar hemangioblastoma: identification of Pro157Leu mutation in the VHL gene.* J Hum Genet, 2000. **45**(1): p. 47-51.

115. Hes, F.J., et al., *Cryptic von Hippel-Lindau disease: germline mutations in patients with haemangioblastoma only.* J Med Genet, 2000. **37**(12): p. 939-943.

116. Neumann, H.P.H., et al., *Germ-line mutations in nonsyndromic pheochromocytoma.* N Engl J Med, 2002. **346**(19): p. 1459-1466.

117. Cascon, A., et al., *Genetics of pheochromocytoma and paraganglioma in Spanish patients.* J Clin Endocrinol Metab, 2009. **94**(5): p. 1701-5.

118. Glenn, G.M., et al., *Von Hippel-Lindau (VHL) disease: distinct phenotypes suggest more than one mutant allele at the VHL locus.* Hum Genet, 1991. **87**(2): p. 207-10.

119. Nordstrom-O'Brien, M., et al., *Genetic analysis of von Hippel-Lindau disease.* Hum Mutat, 2010. **31**(5): p. 521-37.

120. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.

121. Mandriota, S.J., et al., *HIF activation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron.* Cancer Cell, 2002. **1**(5): p. 459-468.

122. Esteban, M.A., et al., *Regulation of E-cadherin expression by VHL and hypoxia-inducible factor.* Cancer Res, 2006. **66**(16585181): p. 3567-3575.

123. Linehan, W.M., R. Srinivasan, and L.S. Schmidt, *The genetic basis of kidney cancer: a metabolic disease.* Nat Rev Urol, 2010. **7**(5): p. 277-85.

124. Zaghloul, N.A. and N. Katsanis, *Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy.* J Clin Invest, 2009. **119**(3): p. 428-37.

125. Maxwell, P.H., et al., *The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis.* Nature, 1999. **399**(6733): p. 271-5.

126. Zhuang, Z., et al., *Somatic HIF2A gain-of-function mutations in paraganglioma with polycythemia.* N Engl J Med, 2012. **367**(10): p. 922-30.

127. Gale, D.P., et al., *Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating HIF2 alpha mutation.* Blood, 2008. **112**(18650473): p. 919-921.

128. Percy, M.J., et al., *A gain-of-function mutation in the HIF2A gene in familial erythrocytosis.* N Engl J Med, 2008. **358**(18184961): p. 162-168.

129. Gordeuk, V.R. and J.T. Prchal, *Vascular complications in Chuvash polycythemia.* Semin Thromb Hemost, 2006. **32**(3): p. 289-94.

130. Brooks, J.T.S., et al., *Variations within oxygen-regulated gene expression in humans.* J Appl Physiol, 2009. **106**(19008490): p. 212-220.

131. Boyer, O., et al., *Mutational analysis of the PLCE1 gene in steroid resistant nephrotic syndrome.* J Med Genet, 2010. **47**(7): p. 445-52.

132. Antonarakis, S.E. and J.S. Beckmann, *Mendelian disorders deserve more attention.* Nat Rev Genet, 2006. **7**(4): p. 277-82.

133. Sidransky, E., *Heterozygosity for a Mendelian disorder as a risk factor for complex disease.* Clin Genet, 2006. **70**(4): p. 275-82.

134. Voskarides, K., et al., *Evidence that NPHS2-R229Q predisposes to proteinuria and renal failure in familial hematuria.* Pediatr Nephrol, 2012. **27**(4): p. 675-9.

135. Caridi, G., et al., *Clinical features and long-term outcome of nephrotic syndrome associated with heterozygous NPHS1 and NPHS2 mutations.* Clin J Am Soc Nephrol, 2009. **4**(6): p. 1065-72.

136. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

137. Gale, D.P., et al., *Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis.* Lancet, 2010. **376**(9743): p. 794-801.

138. Lupski, J.R. and P. Stankiewicz, *Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes.* PLoS Genet, 2005. **1**(6): p. e49.

139. Hollox, E.J., et al., *Psoriasis is associated with increased beta-defensin genomic copy number.* Nat Genet, 2008. **40**(1): p. 23-5.

140. Aitman, T.J., et al., *Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans.* Nature, 2006. **439**(7078): p. 851-5.

141. Copeland, W.C., *Inherited mitochondrial diseases of DNA replication.* Annu Rev Med, 2008. **59**: p. 131-46.

142. Finsterer, J., *Overview on visceral manifestations of mitochondrial disorders.* Neth J Med, 2006. **64**(3): p. 61-71.

143. Chinnery, P.F., et al., *Epigenetics, epidemiology and mitochondrial DNA diseases.* Int J Epidemiol, 2012. **41**(1): p. 177-87.

144. de Borst, M.H., A. Benigni, and G. Remuzzi, *Primer: strategies for identifying genes involved in renal disease.* Nat Clin Pract Nephrol, 2008. **4**(5): p. 265-76.

145. Oudit, G.Y., et al., *Loss of angiotensin-converting enzyme-2 leads to the late development of angiotensin II-dependent glomerulosclerosis.* Am J Pathol, 2006. **168**(6): p. 1808-20.

146. Yoshikawa, M., et al., *Inhibition of histone deacetylase activity suppresses epithelial-to-mesenchymal transition induced by TGF-beta1 in human renal epithelial cells.* J Am Soc Nephrol, 2007. **18**(1): p. 58-65.

147. Shen, B., et al., *The Bradykinin B2 receptor gene is a target of angiotensin II type 1 receptor signaling.* J Am Soc Nephrol, 2007. **18**(4): p. 1140-9.

148. Census. *The Final results of TRNC General Population and Housing Unit Census*. 2006; Available from: http://nufussayimi.devplan.org/Kesin-sonuc-index_en.html.

149. Nephrology, T.S.o., *Registry of the Nephrology, Dialysis and Transplantation in Turkey*, in *Annual Report Books*2008: Istanbul.

150. ONS. *Mid Year population estimates 2008: 13/05/10*. 2011; Available from: http://www.statistics.gov.uk/statbase/Product.asp?vlnk=15106.

151. Institute, T.S. *Midyear Population Projections by Age Groups and Sex*. TurkStat 2008; Available from: http://www.turkstat.gov.tr/VeriBilgi.do?alt_id=39.

152. Ye, S., et al., *An efficient procedure for genotyping single nucleotide polymorphisms.* Nucleic Acids Res, 2001. **29**(17): p. E88-8.

153. Hoffmann, K. and T.H. Lindner, *easyLINKAGE-Plus--automated linkage analyses using large-scale SNP data.* Bioinformatics, 2005. **21**(17): p. 3565-7.

154. Markianos, K., M.J. Daly, and L. Kruglyak, *Efficient multipoint linkage analysis through reduction of inheritance space.* Am J Hum Genet, 2001. **68**(4): p. 963-77.

155. Sobel, E. and K. Lange, *Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics.* Am J Hum Genet, 1996. **58**(6): p. 1323-37.

156. Thiele, H. and P. Nurnberg, *HaploPainter: a tool for drawing pedigrees with complex haplotypes.* Bioinformatics, 2005. **21**(8): p. 1730-2.

157. O'Connell, J.R. and D.E. Weeks, *PedCheck: a program for identification of genotype incompatibilities in linkage analysis.* Am J Hum Genet, 1998. **63**(1): p. 259-66.

158. Schwarz, J.M., et al., *MutationTaster evaluates disease-causing potential of sequence alterations.* Nat Methods, 2010. **7**(8): p. 575-6.

159. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic Acids Res, 2003. **31**(13): p. 3812-4.

160. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

161. Zhou, T., et al., *Generation of induced pluripotent stem cells from urine.* J Am Soc Nephrol, 2011. **22**(7): p. 1221-8.

162. Krieg, M., et al., *Up-regulation of hypoxia-inducible factors HIF-1alpha and HIF-2alpha under normoxic conditions in renal carcinoma cells by von Hippel-Lindau tumor suppressor gene loss of function.* Oncogene, 2000. **19**(48): p. 5435-5443.

163. Eurostat. *European Comission Population Statistics*. 2010; Available from: http://epp.eurostat.ec.europa.eu/portal/page/portal/population/data/main_tables.

164. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR.* Nucleic Acids Res, 2001. **29**(9): p. e45.

165. Agodoa, L. and P. Eggers, *Racial and ethnic disparities in end-stage kidney failure-survival paradoxes in African-Americans.* Semin Dial, 2007. **20**(6): p. 577-85.

166. Palmer Alves, T. and J. Lewis, *Racial differences in chronic kidney disease (CKD) and end-stage renal disease (ESRD) in the United States: a social and economic dilemma.* Clin Nephrol, 2010. **74 Suppl 1**: p. S72-7.

167. Novembre, J., et al., *Genes mirror geography within Europe.* Nature, 2008. **456**(7218): p. 98-101.

168. Irwin, J., et al., *Mitochondrial control region sequences from northern Greece and Greek Cypriots.* Int J Legal Med, 2008. **122**(1): p. 87-9.

169. Baysal, E., et al., *The beta-thalassaemia mutations in the population of Cyprus.* Br J Haematol, 1992. **81**(4): p. 607-9.

170. Baysal, E., et al., *alpha-Thalassaemia in the population of Cyprus.* Br J Haematol, 1995. **89**(3): p. 496-9.

171. Voskarides, K., et al., *COL4A3/COL4A4 mutations producing focal segmental glomerulosclerosis and renal failure in thin basement membrane nephropathy.* J Am Soc Nephrol, 2007. **18**(11): p. 3004-16.

172. Stavrou, C., et al., *Autosomal-dominant medullary cystic kidney disease type 1: clinical and molecular findings in six large Cypriot families.* Kidney Int, 2002. **62**(4): p. 1385-94.

173. Voskarides, K., et al., *COL4A3 founder mutations in Greek-Cypriot families with thin basement membrane nephropathy and focal segmental glomerulosclerosis dating from around 18th century.* Genet Test, 2008. **12**(2): p. 273-8.

174. Deltas, C.C., et al., *Familial Mediterranean fever (FMF) mutations occur frequently in the Greek-Cypriot population of Cyprus.* Genet Test, 2002. **6**(1): p. 15-21.

175. Arcos-Burgos, M. and M. Muenke, *Genetics of population isolates.* Clin Genet, 2002. **61**(4): p. 233-47.

176. Connor, T.M., et al., *Incidence of end-stage renal disease in the Turkish-cypriot population of northern cyprus: a population based study.* PLoS One, 2013. **8**(1): p. e54394.

177. Rosansky, S.J., et al., *Initiation of dialysis at higher GFRs: is the apparent rising tide of early dialysis harmful or helpful?* Kidney Int, 2009. **76**(3): p. 257-61.

178. van Dijk, P.C., et al., *Effect of general population mortality on the north-south mortality gradient in patients on replacement therapy in Europe.* Kidney Int, 2007. **71**(1): p. 53-9.

179. Neild, G.H., D.D. Oygar, and M.B. Hmida, *Can we improve the diagnosis of renal failure? A revised coding system for the Middle East and North Africa.* Saudi J Kidney Dis Transpl, 2011. **22**(4): p. 651-61.

180. Stewart, J.H., et al., *Trends in incidence of treated end-stage renal disease, overall and by primary renal disease, in persons aged 20-64 years in Europe, Canada and the Asia-Pacific region, 1998-2002.* Nephrology (Carlton), 2007. **12**(5): p. 520-7.

181. White, S.L., et al., *How can we achieve global equity in provision of renal replacement therapy?* Bull World Health Organ, 2008. **86**(3): p. 229-37.

182. Shaheen, F.A. and A.A. Al-Khader, *Preventive strategies of renal failure in the Arab world.* Kidney Int Suppl, 2005(98): p. S37-40.

183. Loizou, T., et al., *An epidemiologic study on the prevalence of diabetes, glucose intolerance, and metabolic syndrome in the adult population of the Republic of Cyprus.* Diabetes Care, 2006. **29**(7): p. 1714-5.

184. Suleymanlar, G., et al., *A population-based survey of Chronic REnal Disease In Turkey--the CREDIT study.* Nephrol Dial Transplant, 2011. **26**(6): p. 1862-71.

185. NHANES. *National Health and Nutrition Examination Survey*. Chronic Disease Prevention and Health Promotion 2010 [cited 2011; Available from: http://www.cdc.gov/chronicdisease/resources/publications/AAG/ddt.htm.

186. Lemley, K.V., *Diabetes and chronic kidney disease: lessons from the Pima Indians.* Pediatr Nephrol, 2008. **23**(11): p. 1933-40.

187. Savva, S.C., et al., *Prevalence of overweight and obesity among 11-year-old children in Cyprus, 1997-2003.* Int J Pediatr Obes, 2008. **3**(3): p. 186-92.

188. Olds, T., et al., *Evidence that the prevalence of childhood overweight is plateauing: data from nine countries.* Int J Pediatr Obes, 2011. **6**(5-6): p. 342-60.

189. Suleymanlar, G., et al., *Trends in renal replacement therapy in Turkey, 1996-2008.* Am J Kidney Dis, 2011. **57**(3): p. 456-65.

190. Perneger, T.V., et al., *Diagnosis of hypertensive end-stage renal disease: effect of patient's race.* Am J Epidemiol, 1995. **141**(1): p. 10-5.

191. Wikipedia, ed. *Northern Cyprus*. 2011.

192. Neild, G.H., *What do we know about chronic renal failure in young adults? I. Primary renal disease.* Pediatr Nephrol, 2009. **24**(10): p. 1913-9.

193. Gale, D.P., *The identification of CFHR5 nephropathy.* J R Soc Med, 2011. **104**(5): p. 186-90.

194. Hanley, J.A. and A. Lippman-Hand, *If nothing goes wrong, is everything all right? Interpreting zero numerators.* JAMA, 1983. **249**(13): p. 1743-5.

195. Genin, E., et al., *Estimating the age of rare disease mutations: the example of Triple-A syndrome.* J Med Genet, 2004. **41**(6): p. 445-9.

196. Labuda, D., E. Zietkiewicz, and M. Labuda, *The genetic clock and the age of the founder effect in growing populations: a lesson from French Canadians and Ashkenazim.* Am J Hum Genet, 1997. **61**(3): p. 768-71.

197. Colombo, R. *Dating Mutations*. eLS, 2007. DOI: 10.1002/9780470015902.a0005462.pub2.

198. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-69.

199. Temme, J., et al., *Incidence of renal failure and nephroprotection by RAAS inhibition in heterozygous carriers of X-chromosomal and autosomal recessive Alport mutations.* Kidney Int, 2012. **81**(8): p. 779-83.

200. Jefferson, J.A., et al., *Autosomal dominant Alport syndrome linked to the type IV collage alpha 3 and alpha 4 genes (COL4A3 and COL4A4).* Nephrol Dial Transplant, 1997. **12**(8): p. 1595-9.

201. Ciccarese, M., et al., *Identification of a new mutation in the alpha4(IV) collagen gene in a family with autosomal dominant Alport syndrome and hypercholesterolaemia.* Nephrol Dial Transplant, 2001. **16**(10): p. 2008-12.

202. Pescucci, C., et al., *Autosomal-dominant Alport syndrome: natural history of a disease due to COL4A3 or COL4A4 gene.* Kidney Int, 2004. **65**(5): p. 1598-603.

203. Zouvani, I., et al., *Incidence of thin basement membrane nephropathy in 990 consecutive renal biopsies examined with electron microscopy.* Ultrastruct Pathol, 2008. **32**(6): p. 221-6.

204. Di Angelantonio, E., et al., *Chronic kidney disease and risk of major cardiovascular disease and non-vascular mortality: prospective population based cohort study.* BMJ, 2010. **341**: p. c4986.

205. Bellis, M.A., et al., *Measuring paternal discrepancy and its public health consequences.* J Epidemiol Community Health, 2005. **59**(9): p. 749-54.

206. Coresh, J., et al., *Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third National Health and Nutrition Examination Survey.* Am J Kidney Dis, 2003. **41**(1): p. 1-12.

207. Snyder, J.J., R.N. Foley, and A.J. Collins, *Prevalence of CKD in the United States: a sensitivity analysis using the National Health and Nutrition Examination Survey (NHANES) 1999-2004.* Am J Kidney Dis, 2009. **53**(2): p. 218-28.

208. Mohr, D.N., et al., *Asymptomatic microhematuria and urologic disease. A population-based study.* JAMA, 1986. **256**(2): p. 224-9.

209. Thompson, I.M., *The evaluation of microscopic hematuria: a population-based study.* J Urol, 1987. **138**(5): p. 1189-90.

210. Christodoulou, K., et al., *Chromosome 1 localization of a gene for autosomal dominant medullary cystic kidney disease.* Hum Mol Genet, 1998. **7**(5): p. 905-11.

211. Drmanac, R., et al., *Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays.* Science, 2010. **327**(5961): p. 78-81.

212. Manolio, T.A., L.D. Brooks, and F.S. Collins, *A HapMap harvest of insights into the genetics of common disease.* J Clin Invest, 2008. **118**(5): p. 1590-605.

213. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

214. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.

215. Ng, P.C., et al., *Genetic variation in an individual human exome.* PLoS Genet, 2008. **4**(8): p. e1000160.

216. Bailey-Wilson, J.E. and A.F. Wilson, *Linkage analysis in the next-generation sequencing era.* Hum Hered, 2011. **72**(4): p. 228-36.

217. Abecasis, G.R., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.

218. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.

219. McLaren, W., et al., *Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor.* Bioinformatics, 2010. **26**(16): p. 2069-70.

220. Ahmed, Z.M., et al., *Gene structure and mutant alleles of PCDH15: nonsyndromic deafness DFNB23 and type 1 Usher syndrome.* Hum Genet, 2008. **124**(3): p. 215-23.

221. Alli, Z., et al., *Characterization of murine Dhx32.* Exp Mol Pathol, 2007. **83**(1): p. 115-8.

222. Alli, Z., et al., *A role for DHX32 in regulating T-cell apoptosis.* Anticancer Res, 2007. **27**(1A): p. 373-7.

223. Abdelhaleem, M., T.H. Sun, and M. Ho, *DHX32 expression suggests a role in lymphocyte differentiation.* Anticancer Res, 2005. **25**(4): p. 2645-8.

224. Minagawa, T., et al., *Identification and characterization of a sac domain-containing phosphoinositide 5-phosphatase.* J Biol Chem, 2001. **276**(25): p. 22011-5.

225. Lowe, M., *Structure and function of the Lowe syndrome protein OCRL1.* Traffic, 2005. **6**(9): p. 711-9.

226. Bielas, S.L., et al., *Mutations in INPP5E, encoding inositol polyphosphate-5-phosphatase E, link phosphatidyl inositol signaling to the ciliopathies.* Nat Genet, 2009. **41**(9): p. 1032-6.

227. Zhu, W., et al., *Inpp5f is a polyphosphoinositide phosphatase that regulates cardiac hypertrophic responsiveness.* Circ Res, 2009. **105**(12): p. 1240-7.

228. Trivedi, C.M., et al., *Hdac2 regulates the cardiac hypertrophic response by modulating Gsk3 beta activity.* Nat Med, 2007. **13**(3): p. 324-31.

229. Dorrenhaus, A., et al., *Cultures of exfoliated epithelial cells from different locations of the human urinary tract and the renal tubular system.* Arch Toxicol, 2000. **74**(10): p. 618-26.

230. Belik, R., et al., *Improvements in culturing exfoliated urothelial cells in vitro from human urine.* J Toxicol Environ Health A, 2008. **71**(13-14): p. 923-9.

231. Carvalho, J.C., et al., *p63, CK7, PAX8 and INI-1: an optimal immunohistochemical panel to distinguish poorly differentiated urothelial cell carcinoma from high-grade tumours of the renal collecting system.* Histopathology, 2012. **60**(4): p. 597-608.

232. Helbert, M.J., S.E. Dauwe, and M.E. De Broe, *Flow cytometric immunodissection of the human distal tubule and cortical collecting duct system.* Kidney Int, 2001. **59**(2): p. 554-64.

233. Otto, E.A., et al., *Mutation analysis of 18 nephronophthisis associated ciliopathy disease genes using a DNA pooling and next generation sequencing strategy.* J Med Genet, 2010.

234. Kirby, A., et al., *Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing.* Nat Genet, 2013.

235. Anderson, S., et al., *Sequence and organization of the human mitochondrial genome.* Nature, 1981. **290**(5806): p. 457-65.

236. Falkenberg, M., N.G. Larsson, and C.M. Gustafsson, *DNA replication and transcription in mammalian mitochondria.* Annu Rev Biochem, 2007. **76**: p. 679-99.

237. Schapira, A.H., *Mitochondrial diseases.* Lancet, 2012. **379**(9828): p. 1825-34.

238. Holt, I.J., *Zen and the art of mitochondrial DNA maintenance.* Trends Genet, 2010. **26**(3): p. 103-9.

239. Lightowlers, R.N., et al., *Mammalian mitochondrial genetics: heredity, heteroplasmy and disease.* Trends Genet, 1997. **13**(11): p. 450-5.

240. Durham, S.E., et al., *Normal levels of wild-type mitochondrial DNA maintain cytochrome c oxidase activity for two pathogenic mitochondrial DNA mutations but not for m.3243A-->G.* Am J Hum Genet, 2007. **81**(1): p. 189-95.

241. Cree, L.M., et al., *A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes.* Nat Genet, 2008. **40**(2): p. 249-54.

242. Wai, T., D. Teoli, and E.A. Shoubridge, *The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes.* Nat Genet, 2008. **40**(12): p. 1484-8.

243. Hall, A.M. and R.J. Unwin, *The not so 'mighty chondrion': emergence of renal diseases due to mitochondrial dysfunction.* Nephron Physiol, 2007. **105**(1): p. p1-10.

244. Litonin, D., et al., *Human mitochondrial transcription revisited: only TFAM and TFB2M are required for transcription of the mitochondrial genes in vitro.* J Biol Chem, 2010. **285**(24): p. 18129-33.

245. Bogenhagen, D.F., E.F. Applegate, and B.K. Yoza, *Identification of a promoter for transcription of the heavy strand of human mtDNA: in vitro transcription and deletion mutagenesis.* Cell, 1984. **36**(4): p. 1105-13.

246. Montoya, J., et al., *Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA.* Proc Natl Acad Sci U S A, 1982. **79**(23): p. 7195-9.

247. Bogenhagen, D. and D.A. Clayton, *Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle.* Cell, 1977. **11**(4): p. 719-27.

248. Gelfand, R. and G. Attardi, *Synthesis and turnover of mitochondrial ribonucleic acid in HeLa cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable.* Mol Cell Biol, 1981. **1**(6): p. 497-511.

249. Zollo, O., V. Tiranti, and N. Sondheimer, *Transcriptional requirements of the distal heavy-strand promoter of mtDNA.* Proc Natl Acad Sci U S A, 2012. **109**(17): p. 6508-12.

250. Fish, J., N. Raule, and G. Attardi, *Discovery of a major D-loop replication origin reveals two modes of human mtDNA synthesis.* Science, 2004. **306**(5704): p. 2098-101.

251. Garrido, N., et al., *Composition and dynamics of human mitochondrial nucleoids.* Mol Biol Cell, 2003. **14**(4): p. 1583-96.

252. Ekstrand, M.I., et al., *Mitochondrial transcription factor A regulates mtDNA copy number in mammals.* Hum Mol Genet, 2004. **13**(9): p. 935-44.

253. Tyynismaa, H., et al., *Twinkle helicase is essential for mtDNA maintenance and regulates mtDNA copy number.* Hum Mol Genet, 2004. **13**(24): p. 3219-27.

254. Moreno-Loshuertos, R., et al., *Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants.* Nat Genet, 2006. **38**(11): p. 1261-8.

255. Gray, M.W., G. Burger, and B.F. Lang, *Mitochondrial evolution.* Science, 1999. **283**(5407): p. 1476-81.

256. Handy, D.E. and J. Loscalzo, *Redox regulation of mitochondrial function.* Antioxid Redox Signal, 2012. **16**(11): p. 1323-67.

257. Smeitink, J., L. van den Heuvel, and S. DiMauro, *The genetics and pathology of oxidative phosphorylation.* Nat Rev Genet, 2001. **2**(5): p. 342-52.

258. Gomez-Duran, A., et al., *Unmasking the causes of multifactorial disorders: OXPHOS differences between mitochondrial haplogroups.* Hum Mol Genet, 2010. **19**(17): p. 3343-53.

259. Stewart, J.D., et al., *POLG mutations cause decreased mitochondrial DNA repopulation rates following induced depletion in human fibroblasts.* Biochim Biophys Acta, 2011. **1812**(3): p. 321-5.

260. Shanske, S., et al., *Varying loads of the mitochondrial DNA A3243G mutation in different tissues: implications for diagnosis.* Am J Med Genet A, 2004. **130A**(2): p. 134-7.

261. t Hart, L.M., et al., *Heteroplasmy levels of a mitochondrial gene mutation associated with diabetes mellitus decrease in leucocyte DNA upon aging.* Hum Mutat, 1996. **7**(3): p. 193-7.

262. Rahman, S., et al., *Decrease of 3243 A-->G mtDNA mutation from blood in MELAS syndrome: a longitudinal study.* Am J Hum Genet, 2001. **68**(1): p. 238-40.

263. McDonnell, M.T., et al., *Noninvasive diagnosis of the 3243A > G mitochondrial DNA mutation using urinary epithelial cells.* Eur J Hum Genet, 2004. **12**(9): p. 778-81.

264. Aguer, C., et al., *Galactose enhances oxidative metabolism and reveals mitochondrial dysfunction in human primary muscle cells.* PLoS One, 2011. **6**(12): p. e28536.

265. Valle, I., et al., *PGC-1alpha regulates the mitochondrial antioxidant defense system in vascular endothelial cells.* Cardiovasc Res, 2005. **66**(3): p. 562-73.

266. King, M.P. and G. Attardi, *Human cells lacking mtDNA: repopulation with exogenous mitochondria by complementation.* Science, 1989. **246**(4929): p. 500-3.

267. Morais, R., et al., *Chick embryo cells rendered respiration-deficient by chloramphenicol and ethidium bromide are auxotrophic for pyrimidines.* Biochem Biophys Res Commun, 1980. **94**(1): p. 71-7.

268. Khan, S.M., R.M. Smigrodzki, and R.H. Swerdlow, *Cell and animal models of mtDNA biology: progress and prospects.* Am J Physiol Cell Physiol, 2007. **292**(2): p. C658-69.

269. Gerard, B., et al., *Uridine preserves the expression of respiratory enzyme deficiencies in cultured fibroblasts.* Eur J Pediatr, 1993. **152**(3): p. 270.

270. Rorbach, J. and M. Minczuk, *The post-transcriptional life of mammalian mitochondrial RNA.* Biochem J, 2012. **444**(3): p. 357-73.

271. Mootha, V.K., et al., *Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria.* Cell, 2003. **115**(5): p. 629-40.

272. Hock, M.B. and A. Kralli, *Transcriptional control of mitochondrial biogenesis and function.* Annu Rev Physiol, 2009. **71**: p. 177-203.

273. Ylikallio, E., et al., *High mitochondrial DNA copy number has detrimental effects in mice.* Hum Mol Genet, 2010. **19**(13): p. 2695-705.

274. Eaton, J.S., et al., *Ataxia-telangiectasia mutated kinase regulates ribonucleotide reductase and mitochondrial homeostasis.* J Clin Invest, 2007. **117**(9): p. 2723-34.

275. Lee, H.C., et al., *Increase of mitochondria and mitochondrial DNA in response to oxidative stress in human cells.* Biochem J, 2000. **348 Pt 2**: p. 425-32.

276. Watanabe, M., et al., *Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation.* Nature, 2006. **439**(7075): p. 484-9.

277. Wu, M., et al., *Multiparameter metabolic analysis reveals a close link between attenuated mitochondrial bioenergetic function and enhanced glycolysis dependency in human tumor cells.* Am J Physiol Cell Physiol, 2007. **292**(1): p. C125-36.

278. Hao, H., G. Manfredi, and C.T. Moraes, *Functional and structural features of a tandem duplication of the human mtDNA promoter region.* Am J Hum Genet, 1997. **60**(6): p. 1363-72.

279. Moraes, C.T., et al., *Replication-competent human mitochondrial DNA lacking the heavy-strand promoter region.* Mol Cell Biol, 1991. **11**(3): p. 1631-7.

280. Pezzini, F., et al., *Involvement of the mitochondrial compartment in human NCL fibroblasts.* Biochem Biophys Res Commun, 2011. **416**(1-2): p. 159-64.

281. Cao, Y., et al., *Distinct early molecular responses to mutations causing vLINCL and JNCL presage ATP synthase subunit C accumulation in cerebellar cells.* PLoS One, 2011. **6**(2): p. e17118.

282. Jolly, R.D., et al., *Mitochondrial dysfunction in the neuronal ceroid-lipofuscinoses (Batten disease).* Neurochem Int, 2002. **40**(6): p. 565-71.

283. Carelli, V., et al., *Respiratory function in cybrid cell lines carrying European mtDNA haplogroups: implications for Leber's hereditary optic neuropathy.* Biochim Biophys Acta, 2002. **1588**(1): p. 7-14.

284. Pacheu-Grau, D., et al., *Mitochondrial antibiograms in personalized medicine.* Hum Mol Genet, 2012.

285. Bustamante, E. and P.L. Pedersen, *High aerobic glycolysis of rat hepatoma cells in culture: role of mitochondrial hexokinase.* Proc Natl Acad Sci U S A, 1977. **74**(9): p. 3735-9.

286. Lee, F.S. and M.J. Percy, *The HIF Pathway and Erythrocytosis.* Annu Rev Pathol, 2010.

287. Bond, J., et al., *Dysregulation of the HIF pathway due to VHL mutation causing severe erythrocytosis and pulmonary arterial hypertension.* Blood, 2011. **117**(13): p. 3699-701.

288. Blankenship, C., et al., *Alternate choice of initiation codon produces a biologically active product of the von Hippel Lindau gene with tumor suppressor activity.* Oncogene, 1999. **18**(8): p. 1529-1535.

289. Iliopoulos, O., M. Ohh, and W.G. Kaelin, *pVHL19 is a biologically active product of the von Hippel-Lindau gene arising from internal translation initiation.* Proc Natl Acad Sci U S A, 1998. **95**(20): p. 11661-11666.

290. Schoenfeld, A., E.J. Davidowitz, and R.D. Burk, *A second major native von Hippel-Lindau gene product, initiated from an internal translation start site, functions as a tumor suppressor.* Proc Natl Acad Sci U S A, 1998. **95**(15): p. 8817-8822.

291. Stebbins, C.E., W.G. Kaelin, and N.P. Pavletich, *Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function.* Science, 1999. **284**(5413): p. 455-461.

292. Kibel, A., et al., *Binding of the von Hippel-Lindau tumor suppressor protein to Elongin B and C.* Science, 1995. **269**(5229): p. 1444-1446.

293. Iwai, K., et al., *Identification of the von Hippel-lindau tumor-suppressor protein as part of an active E3 ubiquitin ligase complex.* Proc Natl Acad Sci U S A, 1999. **96**(22): p. 12436-12441.

294. Hon, W.-C., et al., *Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL.* Nature, 2002. **417**(6892): p. 975-978.

295. Ivan, M., et al., *HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing.* Science, 2001. **292**(5516): p. 464-468.

296. Jaakkola, P., et al., *Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation.* Science, 2001. **292**(5516): p. 468-72.

297. Masson, N., et al., *Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation.* Embo J, 2001. **20**(18): p. 5197-206.

298. Bruick, R.K. and S.L. McKnight, *A conserved family of prolyl-4-hydroxylases that modify HIF.* Science, 2001. **294**(5545): p. 1337-40.

299. Epstein, A.C., et al., *C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation.* Cell, 2001. **107**(1): p. 43-54.

300. Paltoglou, S. and B.J. Roberts, *HIF-1alpha and EPAS ubiquitination mediated by the VHL tumour suppressor involves flexibility in the ubiquitination mechanism, similar to other RING E3 ligases.* Oncogene, 2007. **26**(4): p. 604-9.

301. Salceda, S. and J. Caro, *Hypoxia-inducible factor 1alpha (HIF-1alpha) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. Its stabilization by hypoxia depends on redox-induced changes.* J Biol Chem, 1997. **272**(36): p. 22642-7.

302. Frew, I.J. and W. Krek, *pVHL: a multipurpose adaptor protein.* Sci Signal, 2008. **1**(24): p. pe30.

303. Kuznetsova, A.V., et al., *von Hippel-Lindau protein binds hyperphosphorylated large subunit of RNA polymerase II through a proline hydroxylation motif and targets it for ubiquitination.* Proc Natl Acad Sci U S A, 2003. **100**(5): p. 2706-2711.

304. Yang, H., et al., *pVHL acts as an adaptor to promote the inhibitory phosphorylation of the NF-kappaB agonist Card9 by CK2.* Mol Cell, 2007. **28**(17936701): p. 15-27.

305. Roe, J.-S., et al., *p53 stabilization and transactivation by a von Hippel-Lindau protein.* Mol Cell, 2006. **22**(3): p. 395-405.

306. Stickle, N.H., et al., *pVHL modification by NEDD8 is required for fibronectin matrix assembly and suppression of tumor development.* Mol Cell Biol, 2004. **24**(8): p. 3251-3261.

307. Kurban, G., et al., *Collagen matrix assembly is driven by the interaction of von Hippel-Lindau tumor suppressor protein with hydroxylated collagen IV alpha 2.* Oncogene, 2008. **27**(7): p. 1004-1012.

308. Ohh, M., et al., *The von Hippel-Lindau tumor suppressor protein is required for proper assembly of an extracellular fibronectin matrix.* Mol Cell, 1998. **1**(7): p. 959-968.

309. Hergovich, A., et al., *Regulation of microtubule stability by the von Hippel-Lindau tumour suppressor protein pVHL.* Nat Cell Biol, 2003. **5**(1): p. 64-70.

310. Semenza, G.L. and G.L. Wang, *A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation.* Mol. Cell. Biol., 1992. **12**(12): p. 5447-5454.

311. Percy, M.J. and F.S. Lee, *Familial erythrocytosis: molecular links to red blood cell control.* Haematologica, 2008. **93**(18591620): p. 963-967.

312. Ang, S.O., et al., *Endemic polycythemia in Russia: mutation in the VHL gene.* Blood Cells Mol Dis, 2002. **28**(11987242): p. 57-62.

313. Cario, H., et al., *Mutations in the von Hippel-Lindau (VHL) tumor suppressor gene and VHL-haplotype analysis in patients with presumable congenital erythrocytosis.* Haematologica, 2005. **90**(1): p. 19-24.

314. Percy, M.J., et al., *Chuvash-type congenital polycythemia in 4 families of Asian and Western European ancestry.* Blood, 2003. **102**(12702509): p. 1097-1099.

315. Pastore, Y.D., et al., *Mutations in the VHL gene in sporadic apparently congenital polycythemia.* Blood, 2003. **101**(4): p. 1591-1595.

316. Pastore, Y., et al., *Mutations of von Hippel-Lindau tumor-suppressor gene and congenital polycythemia.* Am J Hum Genet, 2003. **73**(2): p. 412-419.

317. Ang, S.O., et al., *Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia.* Nat Genet, 2002. **32**(4): p. 614-621.

318. Perrotta, S., et al., *Von Hippel-Lindau-dependent polycythemia is endemic on the island of Ischia: identification of a novel cluster.* Blood, 2006. **107**(16210343): p. 514-519.

319. Gordeuk, V.R., et al., *Congenital disorder of oxygen sensing: association of the homozygous Chuvash polycythemia VHL mutation with thrombosis and vascular abnormalities but not tumors.* Blood, 2004. **103**(10): p. 3924-3932.

320. Bento, M.C., et al., *Congenital polycythemia with homozygous and heterozygous mutations of von Hippel-Lindau gene: five new Caucasian patients.* Haematologica, 2005. **90**(15642680): p. 128-129.

321. Randi, M.L., et al., *Low frequency of VHL gene mutations in young individuals with polycythemia and high serum erythropoietin.* Haematologica, 2005. **90**(15921386): p. 689-691.

322. Percy, M.J., 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**(3): p. 654-659.

323. Ladroue, C., et al., *PHD2 mutation and congenital erythrocytosis with paraganglioma.* N Engl J Med, 2008. **359**(19092153): p. 2685-2692.

324. Al-Sheikh, M., et al., *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, 2008. **40**(17933562): p. 160-165.

325. Percy, M.J., et al., *A novel erythrocytosis-associated PHD2 mutation suggests the location of a HIF binding groove.* Blood, 2007. **110**(17579185): p. 2193-2196.

326. Takeda, K., et al., *Regulation of adult erythropoiesis by prolyl hydroxylase domain proteins.* Blood, 2008. **111**(18056838): p. 3229-3235.

327. Minamishima, Y.A., et al., *A feedback loop involving the Phd3 prolyl hydroxylase tunes the mammalian hypoxic response in vivo.* Mol Cell Biol, 2009. **29**(19720742): p. 5729-5741.

328. Percy, M.J., et al., *Novel exon 12 mutations in the HIF2A gene associated with erythrocytosis.* Blood, 2008. **111**(18378852): p. 5400-5402.

329. Furlow, P.W., et al., *Erythrocytosis-associated HIF-2alpha mutations demonstrate a critical role for residues C-terminal to the hydroxylacceptor proline.* J Biol Chem, 2009. **284**(19208626): p. 9050-9058.

330. Scortegagna, M., et al., *HIF-2alpha regulates murine hematopoietic development in an erythropoietin-dependent manner.* Blood, 2005. **105**(15626745): p. 3133-3140.

331. Gruber, M., et al., *Acute postnatal ablation of Hif-2alpha results in anemia.* Proc Natl Acad Sci U S A, 2007. **104**(17284606): p. 2301-2306.

332. Martini, M., et al., *A novel heterozygous HIF2AM535I mutation reinforces the role of oxygen sensing pathway disturbances in the pathogenesis of familial erythrocytosis.* Haematologica, 2008. **93**(18508787): p. 1068-1071.

333. Smith, T.G., et al., *Mutation of von Hippel-Lindau tumour suppressor and human cardiopulmonary physiology.* PLoS Med, 2006. **3**(16768548).

334. Smith, T.G., et al., *Mutation of the von Hippel-Lindau gene alters human cardiopulmonary physiology.* Adv Exp Med Biol, 2008. **605**: p. 51-6.

335. Hickey, M.M., et al., *The von Hippel-Lindau Chuvash mutation promotes pulmonary hypertension and fibrosis in mice.* J Clin Invest, 2010. **120**(3): p. 827-39.

336. Wiesener, M.S., et al., *Induction of endothelial PAS domain protein-1 by hypoxia: characterization and comparison with hypoxia-inducible factor-1alpha.* Blood, 1998. **92**(7): p. 2260-8.

337. Sugden, B. and W. Mark, *Clonal transformation of adult human leukocytes by Epstein-Barr virus.* J Virol, 1977. **23**(3): p. 503-8.

338. Nilsson, K. and G. Klein, *Phenotypic and cytogenetic characteristics of human B-lymphoid cell lines and their relevance for the etiology of Burkitt's lymphoma.* Adv Cancer Res, 1982. **37**: p. 319-80.

339. Bornkamm, G.W. and W. Hammerschmidt, *Molecular virology of Epstein-Barr virus.* Philos Trans R Soc Lond B Biol Sci, 2001. **356**(1408): p. 437-59.

340. Schoenfeld, A.R., E.J. Davidowitz, and R.D. Burk, *Elongin BC complex prevents degradation of von Hippel-Lindau tumor suppressor gene products.* Proc Natl Acad Sci U S A, 2000. **97**(15): p. 8507-8512.

341. Kim, J.W., et al., *HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia.* 2006. **3**(3): p. 177-185.

342. Bartrons, R. and J. Caro, *Hypoxia, glucose metabolism and the Warburg's effect.* J Bioenerg Biomembr, 2007. **39**(3): p. 223-9.

343. Semenza, G.L., *Oxygen-dependent regulation of mitochondrial respiration by hypoxia-inducible factor 1.* 2007. **405**(1): p. 1-9.

344. Thomas, G.V., et al., *Hypoxia-inducible factor determines sensitivity to inhibitors of mTOR in kidney cancer.* Nat Med, 2006. **12**(1): p. 122-127.

345. Finley, D., *Recognition and processing of ubiquitin-protein conjugates by the proteasome.* Annu Rev Biochem, 2009. **78**: p. 477-513.

346. Thrower, J.S., et al., *Recognition of the polyubiquitin proteolytic signal.* Embo J, 2000. **19**(1): p. 94-102.

347. Ikeda, F. and I. Dikic, *Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series.* EMBO Rep, 2008. **9**(6): p. 536-42.

348. Kong, X., et al., *Constitutive/hypoxic degradation of HIF-alpha proteins by the proteasome is independent of von Hippel Lindau protein ubiquitylation and the transactivation activity of the protein.* J Biol Chem, 2007. **282**(21): p. 15498-505.

349. Andre, H. and T.S. Pereira, *Identification of an alternative mechanism of degradation of the hypoxia-inducible factor-1alpha.* J Biol Chem, 2008. **283**(43): p. 29375-84.

350. Liu, Y.V., et al., *RACK1 competes with HSP90 for binding to HIF-1alpha and is required for O(2)-independent and HSP90 inhibitor-induced degradation of HIF-1alpha.* Mol Cell, 2007. **25**(2): p. 207-17.

351. Kaluzova, M., et al., *DNA damage is a prerequisite for p53-mediated proteasomal degradation of HIF-1alpha in hypoxic cells and downregulation of the hypoxia marker carbonic anhydrase IX.* Mol Cell Biol, 2004. **24**(13): p. 5757-66.

352. Sdek, P., et al., *MDM2 promotes proteasome-dependent ubiquitin-independent degradation of retinoblastoma protein.* Mol Cell, 2005. **20**(5): p. 699-708.

353. van de Sluis, B., et al., *COMMD1 Promotes pVHL and O2-Independent Proteolysis of HIF-1alpha via HSP90/70.* PLoS One, 2009. **4**(10): p. e7332.

354. Kulikov, R., et al., *Mdm2 facilitates the association of p53 with the proteasome.* Proc Natl Acad Sci U S A, 2010. **107**(22): p. 10038-43.

355. Cho, S., et al., *Binding and regulation of HIF-1alpha by a subunit of the proteasome complex, PSMA7.* FEBS Lett, 2001. **498**(1): p. 62-6.

356. Touitou, R., et al., *A degradation signal located in the C-terminus of p21WAF1/CIP1 is a binding site for the C8 alpha-subunit of the 20S proteasome.* Embo J, 2001. **20**(10): p. 2367-75.

357. Touitou, R., et al., *Epstein-Barr virus EBNA3 proteins bind to the C8/alpha7 subunit of the 20S proteasome and are degraded by 20S proteasomes in vitro, but are very stable in latently infected B cells.* J Gen Virol, 2005. **86**(Pt 5): p. 1269-77.

358. Clifford, S.C., et al., *Contrasting effects on HIF-1alpha regulation by disease-causing pVHL mutations correlate with patterns of tumourigenesis in von Hippel-Lindau disease.* Hum Mol Genet, 2001. **10**(10): p. 1029-1038.

359. Lee, S., et al., *Neuronal apoptosis linked to EglN3 prolyl hydroxylase and familial pheochromocytoma genes: developmental culling and cancer.* Cancer Cell, 2005. **8**(2): p. 155-167.

360. Maxmen, A., *Exome sequencing deciphers rare diseases.* Cell, 2011. **144**(5): p. 635-7.

361. Group, H.G.S., *Building on our inheritance: Genomic technology on heathcare*, D.o. Health, Editor 2012, HM Government. p. 83.

362. Brewer, G.J., *Drug development for orphan diseases in the context of personalized medicine.* Transl Res, 2009. **154**(6): p. 314-22.

363. Kaye, J., et al., *Ethical implications of the use of whole genome methods in medical research.* Eur J Hum Genet, 2010. **18**(4): p. 398-403.

364. Lowrance, W.W. and F.S. Collins, *Ethics. Identifiability in genomic research.* Science, 2007. **317**(5838): p. 600-2.

365. Caulfield, T., et al., *Research ethics recommendations for whole-genome research: consensus statement.* PLoS Biol, 2008. **6**(3): p. e73.

366. Wolf, S.M., et al., *Managing incidental findings in human subjects research: analysis and recommendations.* J Law Med Ethics, 2008. **36**(2): p. 219-48, 211.

367. Church, G.M., Y. Gao, and S. Kosuri, *Next-generation digital information storage in DNA.* Science, 2012. **337**(6102): p. 1628.

368. Choi, M., et al., *Genetic diagnosis by whole exome capture and massively parallel DNA sequencing.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19096-101.

369. Worthey, E.A., et al., *Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease.* Genet Med, 2011. **13**(3): p. 255-62.

Chapter 9 Appendix A

Consent forms and study protocol

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 ☐

I agree to the use of my archived specimens (eg kidney biopsy tissue) for this study, if surplus to clinical requirements ☐

I agree to the creation of stem cells from my urine sample 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).....……..

Protocol version 3.0 for
Genetic Investigation of Inherited Kidney Disease (GIKD)

1. Recruitment
   a. Patient with strong family history of renal disease identified by nephrologist responsible for their care
   b. Inclusion criteria:
      i. Individuals from a family in which two or more individuals are affected by a rare kidney disease (eg glomerulonephritis)
      ii. Individuals from a family in which three or more individuals are affected by a common kidney disease (eg diabetic nephrosclerosis)
   c. Exclusion criteria:
      i. Individuals with an already well−understood genetic renal disease (such as typical adult polycystic kidney disease)
      ii. Families where the inheritance is not clearly Mendelian genetic (ie a single-gene defect is unlikely)
      iii. Individuals who are unable to give appropriate consent
   d. With patient's consent, referral made to GIKD research doctor nephrologist
   e. Patient invited to attend with affected/non-affected family members

2. Acquisition of samples
   a. With consent, blood sample taken from all family members for DNA extraction, analysis and storage
   b. With consent of affected family members, contact made with potentially affected family members to arrange clinical screening (by their GP) for kidney disease and DNA extraction for analysis/storage

3. Genetic Analysis
   a. If clinical features suggest candidate genes responsible:
      i. Known mutations in these genes excluded by sequencing
      ii. Limited SNP analysis of candidate genes to identify SNPs co-segregating with condition (and therefore affected gene/region of gene)
      iii. Appropriate regions will be sequenced
   b. If no clear candidate genes, or if candidate genes excluded, then a genome-wide SNP screen will be done to identify region of genome likely to be responsible
   c. This region will be investigated as above

4. Use of archived tissue
   a. We will review biopsy material from affected individuals in the family in order to determine an accurate clinical phenotype.
      i. This may include archived tissue from family members who are now deceased.
   b. In rare instances we may use historical patient samples that are surplus to clinical requirement for the investigation:
      i. With consent, we may use the tissue to look at gene expression to refine the results of a genome-wide SNP screen.
      ii. It is possible to get limited genetic information from the archived tissue of deceased individuals which can inform the results of a

SNP screen and confirm the results of candidate gene sequencing.

5. Storage of Samples and Information
   a. Clinical details will be anonymised and stored on a secure computer in Professor Maxwell's laboratory. The key for the anonymised patient details will be stored in a locked metal cabinet
   b. DNA will be stored in Professor Maxwell's laboratory for future investigation by us and other researchers in kidney disease
      i. Access to laboratory is swipe-card controlled
      ii. Samples will be labelled with a code
      iii. Key to code will be kept on password-protected computer and in a locked metal cabinet
      iv. Researchers will have access to code to identify samples as necessary

6. Generation of induced Stem Cells from Urine
   a. With consent, urine samples will be collected from up to 3 individuals with kidney disease per family and a corresponding number of healthy relatives.
   b. Renal tubular cells will be isolated from these urine samples and cultured to produce induced pluripotential stem cells.
   c. The stem cells will be used to generate differentiated cell lineages with which to study the effect of the causative mutation in that family.
   d. All cell lines generated by these experiments will be stored in Professor Maxwell's laboratory for the duration of the study.
   e. If any patient decides to withdraw from the study we will discard any cells they have contributed along with their DNA.

7. Further Investigation
   a. Once the causative mutation has been identified the molecular effects of this will be investigated using cell-free and cell-based techniques, as appropriate

8. Dissemination of Results
   a. Anonymised results of each family investigation will be published in peer-reviewed journals
   b. All participants will be informed of outcome of investigation and offered a further clinic appointment with a clinical geneticist to discuss implications of their condition for them and their family

CONFIDENTIAL PATIENT INFORMATION SHEET

Study organiser (Cyprus): Dr Deren Oygar, Consultant Nephrologist, State Hospital, Nicosia.
Study organiser (London): Professor Guy Neild (Professor of Nephrology)

Investigation of familial renal disease in Turkish Cypriot kindreds.

(To describe the symptoms and find the responsible gene for inherited kidney disease in Cyprus)

You are invited to take part in this research study.

Inherited kidney disease is common in Cyprus. We are trying to find the gene responsible for these diseases and make detailed investigation of the urine.

From our urine research we hope to find a particular chemical or protein in the urine which will allow us to identify affected members of the family long before their kidney function starts to deteriorate. Early detection will we hope will lead to new ways to treat the disease.

We will take a small blood sample (10 ml); and collect a sample of urine passed at the beginning of the day. DNA will be made from the blood and stored until use. The urine will be kept frozen. The samples will be taken to the UK for research investigations.

If you have any questions about this study at any time, Please contact **Dr Deren Oygar or Professor Guy Neild**, the study organisers.

Please note you do not have to take part in this study if you do not want to. If you decide not to take part you may withdraw at any time without having to give a reason. Your decision whether to take part will not affect your care and management in any way.

All proposals for research using human subjects are reviewed by an ethics committee before they can proceed. This proposal was reviewed by the Ethics Committee at State Hospital, Nicosia and the Joint UCL/UCLH Committees on the Ethics of Human Research.

Dr Deren Oygar, MD.
Professor GH Neild, MD, FRCP, FRCPath.
Study Organisers
November 2008

CONFIDENTIAL CONSENT FORM FOR PATIENTS

Investigation of familial renal disease in Turkish Cypriot kindreds.

Please read this form and then answer each question as appropriate.

Have you read the information sheet about this study? YES/NO

Have you had an opportunity to ask questions and discuss this study? YES/NO

Have you received satisfactory answers to all your questions? YES/NO

Have you received enough information about this study? YES/NO

Which doctor have you spoken to about this study?……

Do you understand that you are free to withdraw from this study
at any time
without giving a reason for withdrawing
without affecting your future medical care YES/NO

Do you agree to take part in this study? YES/NO

I agree to take part in the Study called "**To describe the clinical phenotype and identify the genotype of familial tubular disease in Turkish Cypriot kindreds"**

I agree to provide a blood sample so that my blood sample can be tested for genes causing kidney disease.

SIGNED……..# DATE…….

NAME (PRINT)**……..# AGE** (IF UNDER 18)**……..**

*Please delete as appropriate
#If participant is under 18, form must be signed by a parent/guardian

SIGNATURE OF INVESTIGATOR…….
DATE……

Chapter 10 Appendix B

Published papers

Appendix B