IDENTIFICATION AND ASSESSMENT OF VARIANTS OF UNCERTAIN SIGNIFICANCE IN FAMILIAL CANCER SYNDROMES

By

ELEANOR CLARE RATTENBERRY

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILIOSOPHY

College of Medicine and Dentistry

University of Birmingham

October 2015

ABSTRACT

The identification of the causative mutation(s) in individuals with familial cancer syndromes informs their clinical management and allows cascade testing of family members, which informs their clinical management in turn. The advent of next generation sequencing (NGS) has revolutionised diagnostic genetic analysis, demonstrated by this thesis. Three novel NGS assays have been developed.

The first two assays allowed more comprehensive analysis of two genetically heterogeneous tumours, phaeochromocytoma/paraganglioma and renal cell carcinoma, by creation of NGS-based gene panel tests. These assays allowed increased detection of germline mutations at a lower cost per gene and reduced processing time compared to previous methods of analysis.

The third assay also uses NGS but, instead, to more thoroughly analyse a single gene. The full gene region for *VHL* was examined at mosaic detection level, with a clinically actionable mutation identified in 18% of patients with von Hippel-Lindau disease in whom a mutation could not be identified by conventional analysis.

The difficulty of providing more comprehensive genetic analysis is the concurrent increase in identification of variants of uncertain significance (VUSs). In depth variant analysis was conducted for all VUSs identified during this research. The reassignment of 17% of these VUSs as pathogenic or benign was enabled.

DEDICATION

To my family for supporting me with this endeavour.

To Caitlin for letting me write this rather than play in the garden. To my mum, for always being there when I need her. To Matt, for everything….

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Eamonn Maher for his advice and support throughout this project. Thanks also to Dr Fiona Macdonald for her supervision of this project.

I wish to thank my colleagues at the West Midlands Regional Genetics Laboratory for supporting my application to undertake this research and their help with this project especially in terms of laboratory assistance provided. Hopefully they got as much from it as I did. In particularly I would like to thank Lindsey Vialard, Kirsten McKay, Anna Yeung, Shaun Green and Vera Cercqueira for their scientific and technical help and support. I would also like to thank Yvonne Wallis for her encouragement and help with this project as head of the Familial Cancer Section of the diagnostic laboratory.

My thanks to the staff and students of Prof. Maher's group at the University of Birmingham for their assistance with my attempt at functional assays.

I would like to thank my funders, the National Institute for Health Research and the VHL alliance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AA	access array
ABI	applied biosystems
ADP	adenosine diphosphate
APS	ammonium persulphate
bHLHZ	basic-helix–loop–helix–zipper
bp	base pair
BSA	bovine serum albumin
ccRCC	clear cell renal cell carcinoma
chRCC	chromophobe renal cell carcinoma
cDNA	complementary DNA (synthesised from RNA)
CDS	coding sequence
CPA	clinical pathology accreditation
CPG	cancer predisposition gene
ddH_2O	double distilled water
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonuclease triphosphate
EDTA	ethylenediaminetetraacetic acid
FFPE	formalin fixed paraffin embedded
FMTC	familial medullary thyroid cancer
H&E	haematoxylin and eosin
hg19	human genome 19
HNPGL	head and neck paraganglioma
IF	incidental finding
kb	kilobase

LOH	loss of heterozygosity
LSDB	locus specific database
MAF	minor allele frequency
Mb	megabases
MEN2	multiple endocrine neoplasia type 2
MMLV	Moloney murine leukaemia virus
mRNA	messenger RNA
NGS	next generation sequencing
OC	oncocytoma
PARP	poly(ADP-ribose) polymerase
PCR	polymerase chain reaction
PPi	pyrophosphate
pRCC	papillary renal cell carcinoma
PTP	picotiter plate
PGL	paraganglioma (abdominal)
PHEO	phaeochromocytoma
pVHL	VHL protein
RCC	renal cell carcinoma
RNA	ribonucleic acid
Rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
TSG	tumour suppressor gene
VHL	von Hippel-Lindau syndrome
VUS	variant of uncertain significance
WMRGL	West Midlands Regional Genetics Laboratory

Chapter 1 Introduction

1.1 A brief introduction to medical genetics

During the last two centuries, understanding of genetics has increased exponentially. From the insights of Gregor Mendel in the 1800s regarding the inheritance of traits in peas via the discovery of the DNA double helix by Watson and Crick in 1953, the field of genetics has exploded. In humans, genetics is now known to be instrumental in everything from eye colour (1) to cancer predisposition (2). Genetics plays a fundamental role in almost every discipline of medicine, clinical genetics is a speciality in its own right and the diagnostic genetics laboratory has become an important part of healthcare systems in developed countries around the world. Figure 1.1 demonstrates some of the landmarks in the study of genetics.

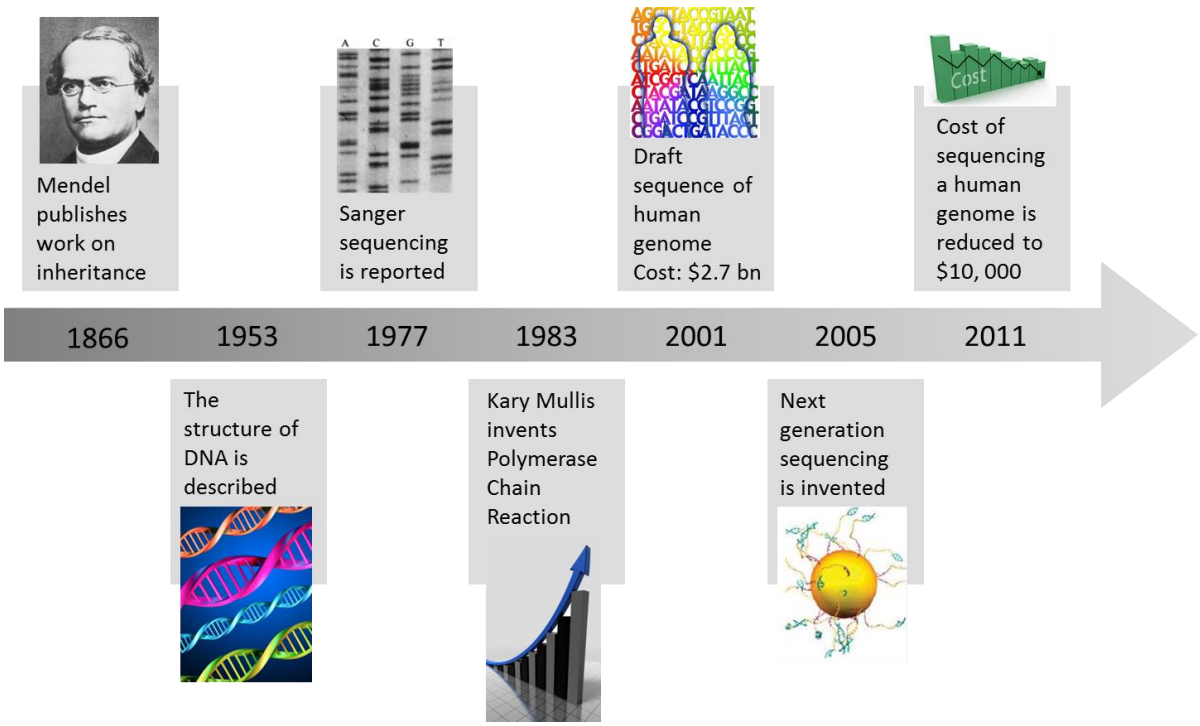


Figure 1.1 Landmarks in genetics
(Adapted from (3)).

Huge leaps forward in the understanding of genetics have often been driven by the advent of new techniques. In 1977 two methods for sequencing the DNA molecule were published (4;5). The Sanger sequencing method was adopted by the genetics community and is still at the heart of diagnostic laboratories' work (6). By 1983 the polymerase chain reaction (PCR) method was published (7), and together with sequencing, it transformed the study of DNA and human genetic disease. PCR and Sanger sequencing on an industrial scale allowed the full sequence of the human genome to be published in 2001 (6). This huge collaborative research effort took 13 years and cost in excess $2.7 billion (6). The sequence was immediately an invaluable tool for the research and diagnostic genetic communities alike. In 2005 genetics was revolutionised again with the commercial launch of the first next generation sequencing (NGS) platform (6). By the time of the commencement of this project in October 2010, enhancement in NGS platforms had resulted in the cost of sequencing a human genome having fallen to under $10, 000, with a time frame of weeks. Vast numbers of genotypes could be generated with comparative ease. Although whole genome or exome sequencing was not yet provided by diagnostic laboratories, NGS technology was starting to be employed in a more targeted manner.

This introduction aims to describe NGS technology, as at the start of this project in 2010, and its role in the diagnostic setting. The application of diagnostic genetics and NGS to cancer will be introduced. This will be followed by a brief introduction to the interpretive challenges presented by the large amounts of data produced by NGS.

1.2 NGS platforms and library preparation methods

At the commencement of this study in 2010, there were three companies providing NGS platforms that were widely commercially available, Roche, Illumina and Applied Biosystems (ABI) (see Table 1.1). All these platforms allowed massively parallel sequencing of clonally amplified DNA molecules that were spatially separated in a flow cell (8).

Table 1.1 Comparison of 3 main NGS platforms
(Adapted from (9)).

Company	Roche	Illumina	ABI
Platform	GS-FLX 454 Genome Sequencer	Genome Analyzer	SOLiD Sequencer
Amplification method	Emulsion PCR	Bridge PCR	Emulsion PCR
Sequencing method	Pyrosequencing	Sequencing by synthesis	Sequencing by ligation
Read length (bp)	~400	~75	~50
Run Time/ Throughput (maximum)	10 hours/ 600Mb	7days/17Gb	7days/15Gb
Raw Accuracy	>99.5%	>98.5%	>99.9%
Cost/Mb	$84.40	$6	$5.80

Additional sequencing platforms were available, albeit not widely purchased, or in development; some of which were reported to be able to sequence unamplified single molecules either synchronously or in real time (10).

Both Illumina and Roche had also invested in smaller scale or 'bench-top' NGS technology, developing the MiSeq and GS Junior respectively. The NGS work performed in the course of this research was carried out on the Roche GS Junior (Chapter 3) and the Illumina MiSeq (Chapter 4). The technology behind these platforms is described in more detail below.

1.2.1 Roche GS Junior

The Roche GS Junior used the same chemistry as the larger Roche GS-FLX but had a smaller output of 70, 000 – 100, 000 reads per run as opposed to >1 million, and a throughput of ~35 Mb per run compared to ~400 Mb. The GS Junior technology worked in exactly the same way as the FLX. A library of template DNA was prepared with adapter oligonucleotides attached. The subsequent clonal amplification, pyrosequencing and image recording processes are summarised in Figure 1.2. The images recorded were analysed for signal-to-noise ratio, filtered according to quality criteria, and subsequently algorithmically translated into a linear sequence output (8;11).

In October 2010, the advantages of using the GS Junior NGS methodology compared to other platforms included longer read lengths and the short run time of only 10 hours (8). The main disadvantage was that the data obtained from homopolymer tracts greater than 4 base pairs in length could be highly unreliable because the luminescence yield was not reliably proportional to the number of bases incorporated (8).

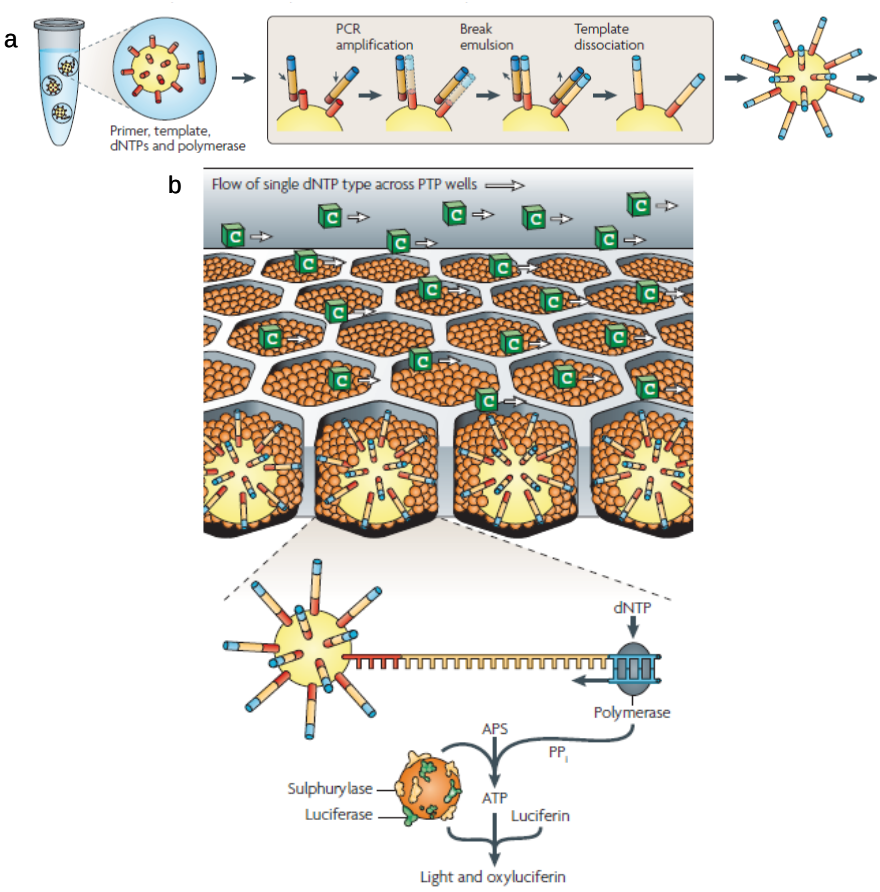


Figure 1.2 Schematic of the Roche NGS workflow

a) Amplification is performed by emulsion PCR whereby single molecules and single beads are separated into individual aqueous microvesicles in an oil-aqueous emulsion. PCR allows clonal expansion of each single molecule, the emulsion is broken, template dissociated and DNA beads enriched (11); b) Sequencing uses a pyrosequencing based methodology. DNA-amplified beads are loaded into individual PicoTiterPlate (PTP) wells with enzyme beads that are coupled with sulphurylase and luciferase. The PTP is loaded onto the GS Junior where it acts as a flow cell. Iterative pyrosequencing is performed by 200 successive flow additions of the 4 dNTPs. In this example, cytosine is shown flowing across the PTP wells. A nucleotide-incorporation event allows pyrophosphate (PPi) release localised to individual PTP wells, which is converted to light photons by the enzyme beads, ATP and luciferin. The light is transmitted through the fibre-optic plate and recorded on a charge-coupled device camera. An image is recorded every time a dNTP is flowed across the PTP (11). (Adapted from (11)).

1.2.2 Illumina MiSeq

Like the Roche GS Junior, the Illumina MiSeq used the same chemistry as its sister machine. Again the capacity was much smaller with a maximum of 1.5Gb of data produced per run compared to up to 200 Gb (8). The main difference was that advances in the technology meant that 151 bp runs were obtainable on the MiSeq before the chemistry was available on the Genome Analyser (later known as the HiSeq), which allowed easier contig assembly (8). The Illumina method required the preparation of a library using template DNA with adapter oligonucleotides attached. This was followed by bridge amplification, sequencing by synthesis and image recording, as shown in Figure 1.4. The images recorded and clusters were identified; then bases were called using signal intensity profiles translated into a linear sequence output with quality criteria (8;11).

The main advantage of using the Illumina NGS methodology, compared to other platforms, was its higher output and lower reagent costs (8;12). The main disadvantage was the higher error rate and the shorter reads compared to the GS Junior (8;12).

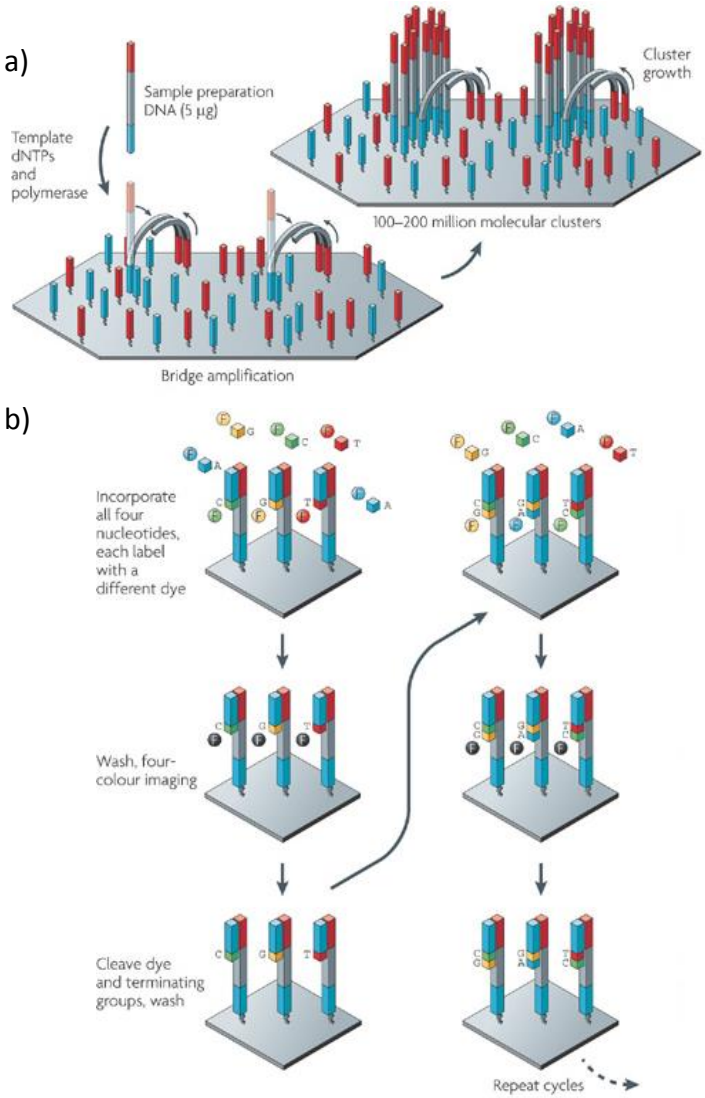


Figure 1.3 Schematic of the Illumina NGS workflow
a) Amplification is performed by bridge PCR, which is a type of solid phase amplification. This process comprises the binding of a single stranded template molecule to a primer immobilised on a flow cell, the extension of the template and its amplification by bridge PCR with further immobilised primers that are in close physical proximity. This creates clusters of identical molecules (11); b) Sequencing by synthesis uses a four-colour cyclic reversible termination method. In each sequencing cycle, all four dNTPs (A,C,G and T), which are fluorescently labelled, are washed over the flow cell. Each dNTP is blocked by a reversible chain terminator which means only a single dNTP can be added to a nucleotide chain in each cycle. An image is then captured and the chain terminator is enzymatically cleaved to allow incorporation of the next dNTP. The clusters are shown as single molecules in the image for illustrative purposes (11). (Adapted from (11)).

1.2.3 NGS target barcoding

If running more than one individual DNA sample on an NGS platform, it is imperative to have a method to differentiate products. Molecular barcodes consist of short runs of oligonucleotides which are included between the target sequence and the platform specific adaptor (13;14). Each DNA is assigned a different barcode, which is incorporated either by ligation or during amplification. Once NGS is completed each read can be bioinformatically assigned to its DNA of origin using the barcode (13;14).

1.2.4 NGS target enrichment methods

If whole genome sequencing is not being performed a method of target enrichment must be carried out prior to sequencing. Since neither the Roche GS Junior, nor the Illumina MiSeq, have the capacity for whole genome sequencing, target enrichment is vital.

Target enrichment can be largely divided into target hybridisation and target amplification methods. Target hybridisation works on the principle of using probes to capture the DNA fragments of interest from a total genomic DNA. It can be performed either in solution or on array and in 2010 commercial examples of this methodology included Agilent SureSelect and Nimblegen. Target amplification methods rely on polymerase chain reaction (PCR) and include simple PCR, long PCR and multiplex PCR. Automated versions of these methods are available and benefit from being much less labour intensive; examples in 2010 included RainDance and Fluidigm. Two different target amplification methods were used during this research. In Chapter 3 the Fluidigm Access Array was used for multiplex PCR and in Chapter 4 long PCR and the Illumina Nextera XT kit were used. The theory behind these

methodologies has been described below. Due to the fast pace of development of all NGS related technology the version of the technology used for this research is described.

1.2.4.1 Fluidigm

The Fluidigm Access Array (AA) was used for target enrichment in Chapter 3. In 2010, the Fluidigm Access Array system allowed the concurrent amplification of 2, 304 PCR reactions (15). The automated system worked using microchannels and valves to create microreactors, which allowed mixing of each of the 48 patient DNAs with each of the 48 primers sets in nanolitre reaction volumes (15). The system also allowed incorporation of adaptor sequences and molecular barcodes (see section 1.2.3) using a four primer 'step-out' PCR (see Figure 1.4), which could also be performed as two independent rounds of PCR if required. The samples and the primers could be different on every set-up. Advantages of this method included its low cost, the low starting concentration of DNA required and high uniformity between amplicons. The main disadvantage was the rigid format, all samples had to be amplified for the same 48 targets and arrays could only be used once which meant the best value for money could only be achieved with a full array, which was not always practical.

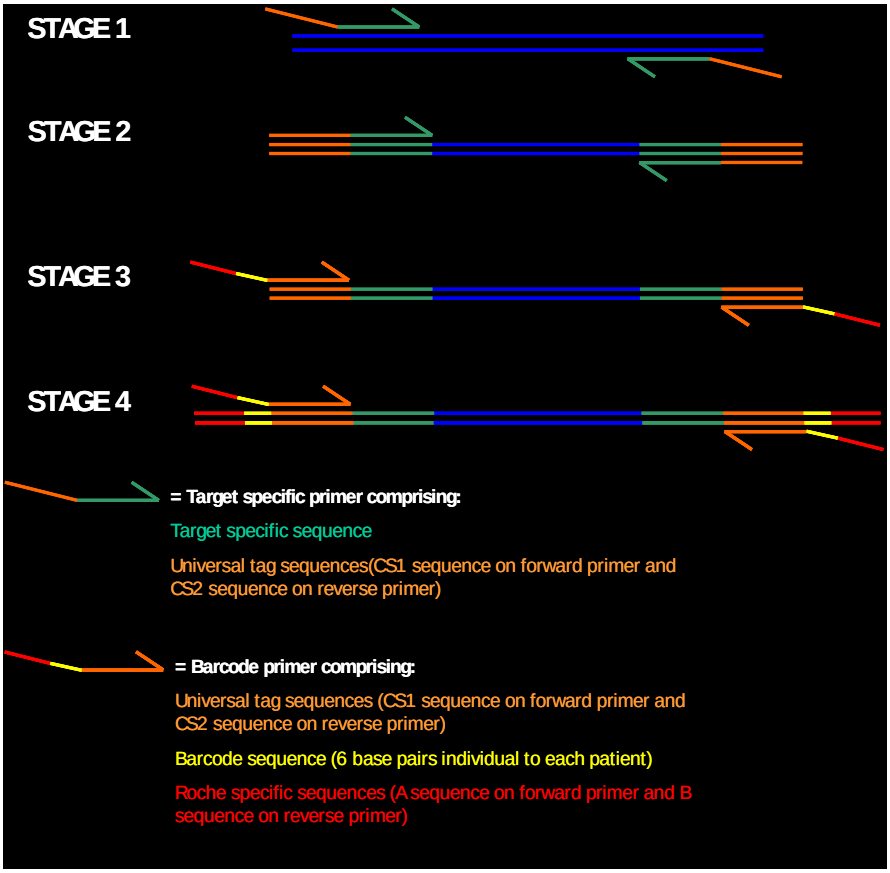


Figure 1.4 Schematic of the Fluidigm four primer system used to allow amplicon-based multi-patient NGS
Stage 1 and 2: the target specific primers bind and amplify the region of interest. These primers are at a low concentration and run out at an early stage of the PCR; Stage 3 and 4: the barcode primers take over the PCR in a 'step-out' reaction. The barcode primers anneal to the universal tag sequences to allow amplification. The final PCR product consists of the Roche specific A and B sequences (red), the individual patient barcode twice (yellow), the CS1 and CS2 universal sequences (orange) and the target region (blue) including target specific primer sequence (green).

1.2.4.2 Long PCR and Illumina Nextera XT

Long PCR and Nextera XT were used for target enrichment in Chapter 4 of the project. Long PCR simply uses a DNA polymerase with both processivity and proof reading abilities to amplify larger regions of DNA. In this project the SequalPrep™ Long PCR Kit was used. The

Illumina Nextera XT system allowed the processing of the long PCR products into barcoded fragmented molecules which included the required adaptor for sequencing on the MiSeq. First, a 'tagmentation' was performed, during which a sample was fragmented and tagged with adapters (16). The fragmentation used a proprietary enzyme named a transposome, which is a derivative of the Tn5 transposase with hyperactivity (17). It cut the DNA and then covalently tagged the 5' end with the Illumina sequencing adapters, as shown in Figure 1.5 (16). Sample indexes and the p5 and p7 sequences, which allowed binding to the flow cell, were then added by a 12 cycle PCR (16). Next, magnetic beads were used to select the DNA molecules at the optimum size of around 300 bp (16). Finally the concentration of the indexed libraries was normalised in order to make sure they were all equally represented (16). Following this the libraries could be pooled and sequenced (16).

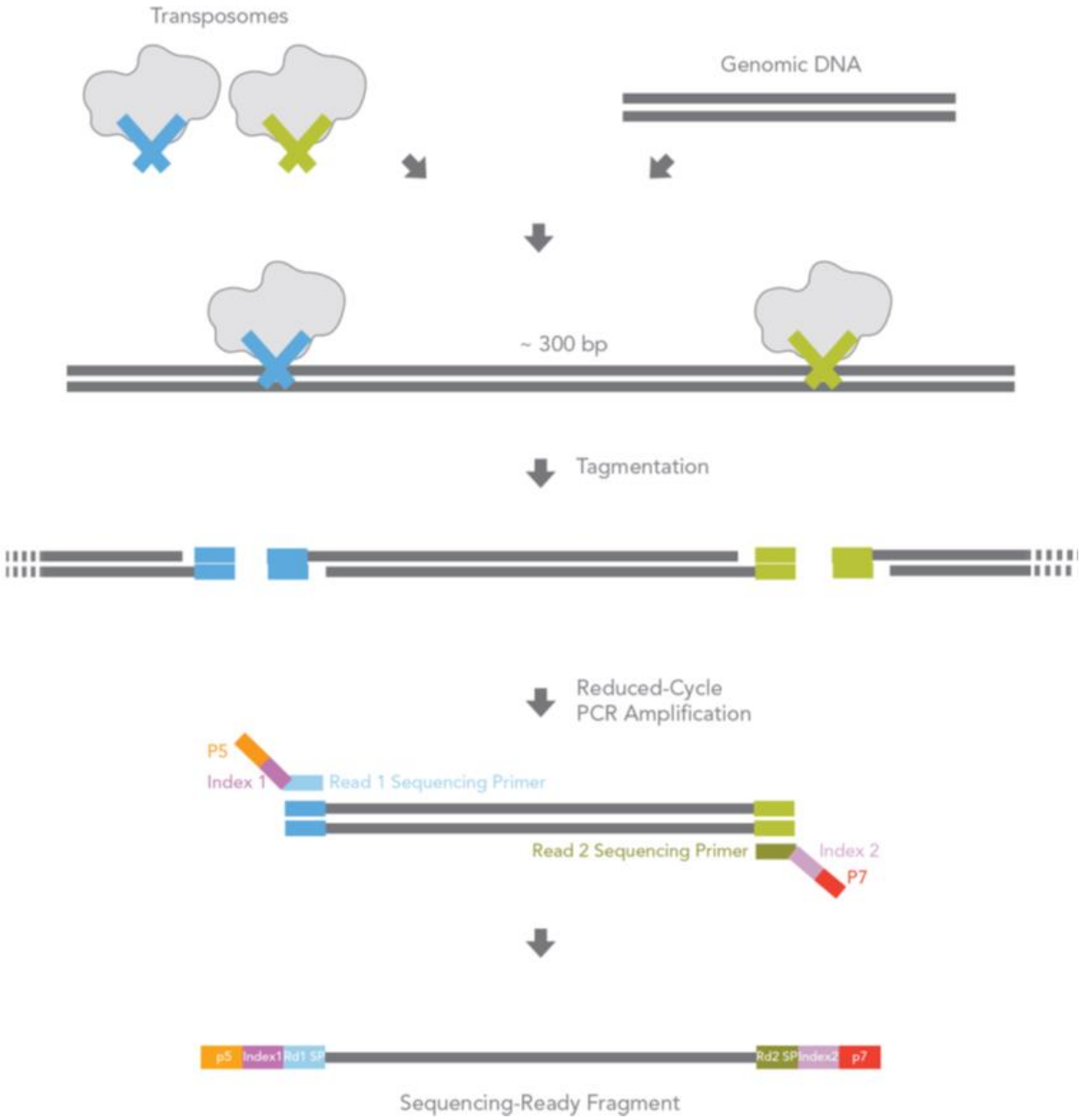


Figure 1.5 Illumina Nextera XT process
DNA is fragmented at an average of 300 bp intervals and tagged with sequencing adaptors (blue and green) using a 'transposome' in a process named tagmentation. A reduced cycle PCR uses tailed primers to bind to the common sequencing primer DNA sequences and add sample specific indexes (purple) with flow cell binding adapters/amplification primers (orange P5; red p7). (Adapted from (16)).

1.2.5 NGS in the diagnostic lab

In 2010 , the field of diagnostic molecular genetics was dominated by low-scale targeted gene and/or mutation analysis (18). In a 2008 paper, ten Bosch and Grody wrote that 'eventually, the perceived clinical benefit of whole-genome sequencing will outweigh the cost of the procedure, allowing for these tests to be performed on a routine basis for diagnostic purposes'. In the meantime, small-scale NGS platforms and target enrichment allowed the implementation of the technology in the diagnostic setting. The main application in development was screening patients who have genetically heterogeneous conditions, such as non-syndromic hearing loss (19). In the short-term, other possible applications included pharmacogenetics, stratified treatment regimens via tumour analysis and metagenomics (9).

1.2.5.1 Advantages and disadvantages of NGS in the diagnostic lab

The use of NGS in diagnostics had, and continues to have, many advantages including the increased capacity and speed but decreased cost per base. This releases the potential for making use of all the information a patient's genome has to provide. The development of more comprehensive screens for genetically heterogeneous conditions could, and indeed does, lead to assays with much greater analytical validity. Fundamentally, more people with a heterogeneous genetic disease will have a pathogenic mutation detected. However, there were also challenges and these challenges remain. Bioinformatic analysis and data storage present a problem but the main long-term stumbling block is that, along with the increased identification of pathogenic mutations, there will automatically be increased identification of variants of unknown significance (VUSs) (see section 1.4).

1.3 Genetic predisposition to cancer

Cancer predisposition genes (CPGs) were defined in a 2014 review by Nazneen Rahman as 'those genes in which rare mutations confer high or moderate risks of cancer (greater than two-fold relative risks) and those for which at least 5% of individuals with the relevant mutations develop cancer' (2). However the majority of CPGs far exceed these criteria for both risk and penetrance (2). In the same literature review, 114 CPGs were identified throughout the genome, using the above definition (2).

1.3.1 Mutational mechanism

CPGs can be subdivided based on the mechanism of action by which mutations within them are pathogenic, either by causing a gain-of-function or a loss-of-function.

1.3.1.1 Gain of function mutations

Gain of function mutations alter a gene product so that its protein has a novel or enhanced function compared to normal. These mutations occur in CPGs called proto-oncogenes when in their normal state and oncogenes once mutated. There are three basic mechanisms of activating a proto-oncogene to become an oncogene: 1) point mutation; 2) gene amplification; 3) chromosome rearrangement (20).

A point mutation can cause activation of an oncogene when it resides within or near the gene and leads to increased activity or loss of regulation of the protein produced as a result of changes in the protein's structure. An example of this is *RET*, a proto-oncogene that causes predisposition to multiple endocrine neoplasia type 2 (MEN2) / familial medullary thyroid cancer (FMTC) when specific amino acids are mutated (21). Mutations of these

particular amino acids results in constitutive activation of the RET protein, a receptor tyrosine kinase (21). This is the only mechanism of proto-oncogene activation which causes both inherited cancer due to mutations occurring in the germline and cancer as a result of somatic mutations.

Gene amplification is a common mechanism for aberrant overexpression of a gene in a tumour. Amplified genes can be extra-chromosomal, forming structures called double minutes, or intra-chromosomal, forming homogeneously staining regions (22). An example is *MYCN*, which is commonly amplified in neuroblastomas (23). *MYCN* encodes a basic-helix–loop–helix–zipper (bHLHZ) transcription factor that plays an important role in the regulation of gene expression associated with a range of cellular processes, including proliferation, growth, apoptosis, energy metabolism, and differentiation (23). Gene amplification occurs somatically.

A chromosome rearrangement is an alteration in which a chromosome, or part of a chromosome, becomes attached to or interchanged with another chromosome or part of a chromosome. The best understood mechanism by which a translocation can activate a proto-oncogene is where a fusion gene is created that puts the control of the expression of the proto-oncogene under the control of another, often constitutively active, gene. The classic example of this is the Philadelphia chromosome, a translocation between the long arms of chromosomes 9 and 22, t(9:22)(q34;q11) (24). This translocation causes the fusion of the *BCR* gene with the *ABL1* gene, which contains a tyrosine kinase domain, and leads to increased levels of tyrosine kinase activity within cells (24). The effects of this increased tyrosine kinase activity include a huge increase in myeloid cell numbers as the result of

increased proliferation or decreased apoptosis of a hematopoietic stem cell or progenitor cell and premature release of immature myeloid cells into the circulation (24). Alternatively, a chromosomal rearrangement can affect the expression of an oncogene by relocating it to a region of active chromatin. This occurs in Burkitt's lymphoma, where the *MYC* gene is moved to an area of active gene expression, most commonly close to the *IGH* gene (25).

1.3.1.2 Loss of function mutations

In comparison, loss of function mutations occur in genes called tumour suppressor genes (TSGs). The mechanism of TSG pathogenesis was first postulated by Alfred Knudson, in an epidemiological study looking at unilateral and bilateral retinoblastoma patients, where he proposed that the two subsets were mechanistically linked (26). He suggested that, in the bilateral form, an inherited mutation provided the first 'hit' and the second 'hit' occurred somatically leading to both copies of the gene being inactivated (26). Whereas, in the sporadic unilateral form, both 'hits' occurred somatically (26). His theory, known as 'Knudson's two hit hypothesis', was demonstrated to be true once the *RB1* gene was cloned in 1986 (27). Knudson's two hit hypothesis explains why, while at a cellular level TSG mutations are recessive because they require inactivation of both copies of the gene for a phenotype to be observed, they have a dominant inheritance pattern, since anyone inheriting the mutation is at risk of disease (27).

1.3.2 Identification and location of CPGs

The majority of CPGs were discovered by genome-wide linkage analysis during the 1990s (2). The advent of next generation sequencing technology has led to additional CPGs being

identified via methodologies such as exome and genome sequencing (2), an example of which is the *MAX* gene (28). The remaining genes were identified by candidate strategies that, although occasionally successful, have been largely unfruitful (2). CPGs are located throughout the genome with little evidence of chromosomal clustering (see Figure 1.6).

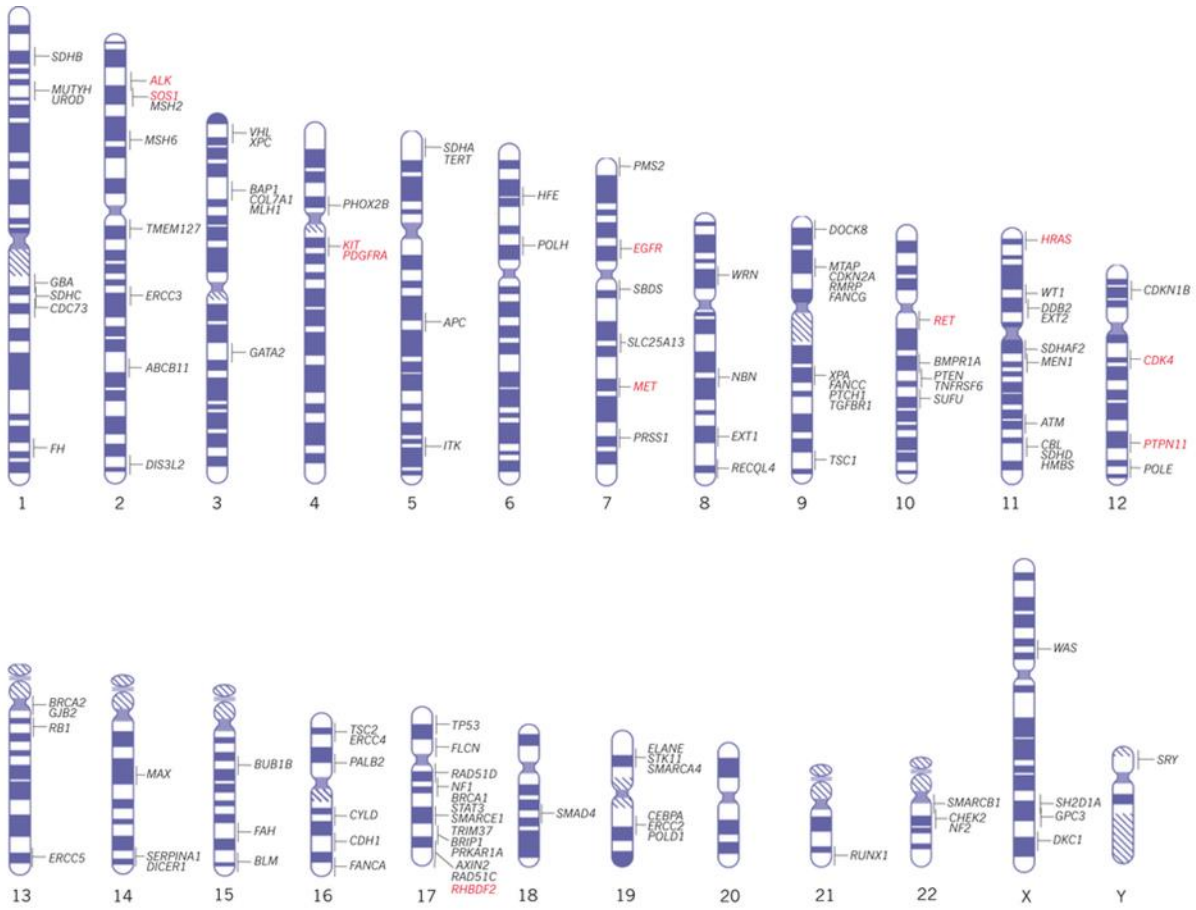


Figure 1.6 Chromosomal locations of cancer predisposition genes as of 2014
Genes which lead to cancer via gain-of-function mutations are in red and those where cancer is caused by loss-of-function mutations in black. (Adapted from (2)).

1.3.3 CPG inheritance patterns

There is no set inheritance pattern for CPGs; of the 114 genes identified in Rahman's review, 65 showed an autosomal dominant inheritance pattern, 28 an autosomal recessive

inheritance pattern, 4 a X-linked inheritance pattern and 1 a Y-linked inheritance pattern (2). An additional 16 genes show both autosomal dominant and recessive inheritance, with a cancer phenotype noted in the case of both monoallelic and biallelic mutations (2). For eight of these genes the recessive disease is more severe than the dominant disease (2). For example, in *BRCA2*, monoallelic mutations cause increased risk of breast and ovarian cancer in adulthood, whereas biallelic mutations cause Fanconi anaemia (29). The identified CPGs were mainly tumour suppressor genes that require inactivation of both alleles for pathogenesis. However, haploinsufficiency and dominant-negative mechanisms also occur. Eleven of the identified CPGs were proto-oncogenes (2).

1.3.4 CPG functions

CPGs have a vast range of functions, many are expressed in all cells and perform essential functions in activities such as cell cycle regulation and DNA repair (2). However, mutations in genes involved in these very general functions can lead to highly specific cancer phenotypes (2). The reasons behind this have been puzzling scientists for decades (2). For example, the *SDHB* gene encodes a subunit of the succinate dehydrogenase complex which participates in both the citric acid cycle and the electron transport chain yet, despite the apparently universal role of the gene, the vast majority of patients with a germline loss of function mutation in this gene only get tumours of the paraganglia (30). In other cases, the normal function of genes relates closely to the cancer phenotype caused by the disruption of that function. For example, loss of function mutations in *ABCB11* can lead to hepatocellular carcinoma. The gene is expressed in the hepatocyte canalicular membrane, where its

normal function is to export primary bile acids against high concentration gradients (31); loss of the normally functioning gene causes hepatic overload and liver cirrhosis, which in turn leads to an increased hepatocellular carcinoma risk (2).

At present, it is estimated that CPG mutations are responsible for about 3% of cancers (2). It has been postulated that this is an underestimate, as there are still CPGs yet to be identified and not all CPG have been fully characterised with regard to the tumours they may cause (2). Whilst some cancers are caused by pathogenic mutations in a single gene, as is the case with retinoblastoma and *RB1* (32), other tumours, such as phaeochromocytoma, can be caused by mutations in many different genes and not all genes may have been identified yet (33). A mutation in a CPG is found in approximately a third of patients with a phaeochromocytoma (33), whereas in patient with bilateral retinoblastoma, an RB1 mutation is identified 95% of cases and the remaining 5% are thought likely to remain undetected due to low-level mosaicism (34).

1.3.5 CPG non-cancer phenotypes

Of the 114 CPGs reviewed by Rahman, over 75% cause clinical features in addition to cancer (2). These features are often very important in the clinical diagnosis of cancer predisposing syndromes since they are more readily distinguishable than the cancer alone (2). However, it is believed that fewer than the recognised 75% of CPGs actually cause additional clinical features and that the current overrepresentation is due to ascertainment bias, as families with additional features are more readily identifiable (2). The additional clinical features

caused by CPGs are diverse and include various skin manifestations, neurological features, skeletal features, dysmorphism and developmental delay (2).

1.3.6 CPG cancer risks

Whilst pathogenic mutations in CPGs increase an individual's risk of getting cancer, the exact nature of the risk is highly complex with different mutations conferring different risks for different cancers (2). Pathogenic mutations in the *BRCA2* gene provide a good example of this. An individual with a loss of function mutation in *BRCA2* has a substantially increased risk of breast and ovarian cancer over their lifetime, but a much smaller increased risk of prostate and pancreatic cancer (35). Additionally different loss of function mutations within *BRCA2* confer different risks, despite the majority of mutations causing a truncated protein which should lead to nonsense mediated decay of the RNA. Those mutations in the central part of the gene give a much higher risk of ovarian cancer as compared to breast cancer, as opposed to mutations in the rest of the gene (36). The reasons for this are not understood.

1.3.7 Diagnostic CPG testing

There are many benefits to performing diagnostic testing on patients who have developed cancer and identifying their pathogenic mutation(s), if present:

1. Psychological benefit - The majority of patients find it useful to know why their cancer occurred (2);

2. Treatment decisions - In terms of management, identifying a mutation in a CPG can help decide if surgery to remove the cancer should be conservative or radical (2). It can also inform the type of therapy which may be instigated; indeed some therapies can be targeted to the specific CPG that is mutated. For example, *BRCA1* or *BRCA2* mutation carriers particularly benefit from the use of Poly(ADP-ribose) polymerase (PARP) inhibitor treatment, which takes advantage of the fact that BRCA1 or BRCA2 dysfunction in combination with the inhibition of the PARP enzyme leads to chromosomal instability, cell cycle arrest and subsequent apoptosis, because BRCA1 and 2 and PARP are normally involved in two distinct DNA repair mechanisms (37);
3. Utility of screening and risk reduction - If a patient is shown to have a CPG mutation the risk of a new primary and/or secondary cancer are often much increased and therefore rigorous screening or risk reducing surgery should be considered, where appropriate. For example, women with a *BRCA1* or *BRCA2* mutation who have breast cancer and have a bilateral mastectomy are less likely to die of breast cancer than those who only have a unilateral mastectomy of the breast which contains the cancer at diagnosis (38);
4. Prognosis – Identification of the responsible CPG provides more accurate survival data. For example, a study in the Netherlands found that patients who met the diagnostic criteria for hereditary diffuse gastric cancer, and in whom a *CDH1* gene mutation was identified, had a shorter survival time than those patients for whom a *CDH1* mutation was not identified (39);
5. Management of associated phenotypes – Identification of a CPG mutation may help manage the CPG related phenotypes that may occur in addition to the cancer. For

example, some *WT1* mutations give rise to renal dysfunction which needs early intervention and monitoring (2);

6. Predictive testing – Following identification of a CPG mutation in an affected individual, predictive testing can then be offered to at-risk family members (as discussed in the following section).

1.3.8 Predictive CPG mutation testing in family members

In addition to helping the affected individual themselves, the identification of a mutation(s) in a CPG allows dissemination of the knowledge of the genetic risk within the family and genetic analysis of at-risk family members, as appropriate, to determine if they are also carrying the pathogenic mutation(s).

1. Positive predictive test - The identification of currently unaffected relatives who have the pathogenic change(s) mean that surveillance and preventative/risk-reducing measures can be implemented as early as possible. The precise measures vary with the cancer. For the majority of cancers, surveillance involves imaging the organ or tissue at risk. However, biomarkers can also be used (2). For example functional paragangliomas and phaeochromocytomas produce catecholamines and the level of these in the blood can be monitored (2). Prevention involves removal of the tissue that is at risk, for example the thyroid for *RET* mutation carriers (40);
2. Negative predictive test - It is equally important to identify family members who do not have the mutation, since the absence of the mutation greatly reduces the individual's risk of getting cancer. Notably, however, this reduction is not always

thought to be back to the level of the general population. For example, in breast cancer families, studies have shown that mutation negative family members may still have a slightly increased risk compared to that of the general population (41).

1.4 Variants of unknown significance (VUSs)

The application of genetics within the clinical context depends upon the identification of clearly pathogenic mutations. When a pathogenic mutation is found it can be very useful for both the management of the patient and the prevention or early detection of cancer in their family members. However, genetic analysis frequently results in the identification of variants of unknown significance (VUSs). This has been a well-recognised problem, for the diagnostic laboratory since the introduction of mutation scanning and continues to be a problem, even in genes which have been extensively examined already. This issue is discussed more comprehensively in Chapter 5.

Chapter 2 General Methods

The general methods described below were used as specified unless described otherwise within individual chapters. In addition, it should be noted that procedures which ultimately led to reporting patient results included appropriate procedures to comply with Clinical Pathology Accreditation (CPA) laboratory standards.

2.1 Nucleic Acid Extraction

2.1.1 DNA extraction from blood

DNA extraction was generously performed by colleagues at the West Midlands Regional Genetics Laboratory (WMRGL). DNA was extracted from lymphocytes using the Gentra Puregene Blood Kit (Qiagen) according to the manufacturer‘s instructions. Extracted DNA was then quantified using a Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). Some DNAs extracted from blood were received from other Regional Genetics Laboratories, where alternative DNA extraction methods were used.

2.1.2 RNA extraction from blood

RNA extraction also was generously performed by colleagues at the West Midlands Regional Genetics Laboratory (WMRGL). Extraction of RNA from whole blood samples was performed using the RNeasy Mini Kit (Qiagen). The process involves two stages, in the first stage erythrocytes are lysed and the leucocytes pelleted, then the pelleted leucocytes are lysed with guanidium-isothiocyanate and homogenized. In the second stage ethanol is added to the lysate to semi-precipitate the nucleic acids present and provide ideal binding conditions.

The lysate is then loaded onto the RNeasy silica membrane, to which RNA binds, and contaminants are efficiently washed through, finally the RNA is eluted in water. The kit was used as per the manufacturer's instructions.

2.1.3 DNA extraction from formalin fixed paraffin embedded (FFPE) tumour tissue

DNA extraction from FFPE tumour tissue was again generously performed by colleagues the West Midlands Regional Genetics Laboratory (WMRGL). Material was received either as scrolls or mounted on slides, with or without a matched haematoxylin and eosin (H&E) stained slide with the tumour marked out.

If mounted slides were received, slides were heated in an oven at 63 ^{o}C for 20 minutes. Each slide was washed in xylene (Sigma), then 100% ethanol (Sigma) and then ddH_2O. The FFPE tissue was then removed to a 2 ml tube containing 150 µl ATL (Qiagen) and 30 µl proteinase K (Qiagen). If a marked H&E slide was available this was used to enable macro-dissection of the tumour tissue.

If scrolls were received, they were placed in a 2 ml tube. 1 ml of xylene (Sigma) was added, the tube vortexed and centrifuged at 13, 000 rpm for 2 minutes, and the liquid removed from the pellet. Then 1 ml of 100% ethanol (Sigma) was added, the tube vortexed and centrifuged at 13, 000 rpm for 5 minutes, and the liquid removed from the pellet. The pellets were then left to air dry. 150 µl ATL (Qiagen) and 30 µl proteinase K (Qiagen) was added to each pellet and the sample left at 56 ^{o}C until it was fully digested.

After this point the rest of the extraction protocol was performed using the QIAcube (Qiagen), according to the manufacturer's instructions using the DNA FFPE tissue protocol.

2.2 cDNA synthesis

The cDNA was prepared from total RNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems). The kit uses the Moloney murine leukaemia virus (MMLV) reverse transcriptase enzyme (branded as MultiScribe™) and random hexamer primers. The kit was used as per the manufacturer's instructions. In brief, a master mix of the required reagents was made up, as per Table 2.1, for the required number of samples and water controls. The master mix was divided into 0.5 µl tubes (17 µl/tube) and 17 µl of RNA added. The tubes were then placed on a Tetrad PTC-225 Thermo Cycler (MJ Research) and underwent the thermal cycling profile given in Table 2.2. Finally the amount of cDNA was measured with a NanoDrop spectrophotometer (Thermo Scientific) and diluted to a working concentration of 200 ng/µl with nuclease free water.

Table 2.1 High Capacity cDNA Reverse Transcription kit master mix reagents and volumes

Component	Volume per reaction (µl)
10x RT Buffer (Applied Biosystems)	3.4
25x dNTP Mix (100 mM; Applied Biosystems)	1.36
10x RT Random Primers (Applied Biosystems)	3.4
MultiScribe™ Reverse Transcriptase (Applied Biosystems)	1.7
RNase Inhibitor (Applied Biosystems)	1.7
Nuclease free water (Applied Biosystems)	5.44
TOTAL	17

Table 2.2 Thermal cycling conditions for High Capacity cDNA Reverse Transcription kit

Step	Temperature	Time
1	25 ^{o}C	10 minutes
2	37 ^{o}C	120 minutes
3	85 ^{o}C	5 seconds
4	4^{o}C	∞

2.3 Polymerase Chain Reaction (PCR)

PCR was used to amplify targeted regions of DNA. Primers were designed using Primer3 (42) and tagged with a common sequence called M13 in order to allow efficient Sanger sequencing. The M13 forward sequence is TGTAAAACGACGGCCAGT and the reverse sequence is CAGGAAACAGCTATGACC. The PCR reaction was made up, as described in Table 2.3. A negative control was run with all PCRs to exclude DNA contamination, in which ddH_20 replaced the DNA. The same ddH_20 was used for any dilutions required during the procedure. The thermal cycling conditions were as shown in Table 2.4. A touchdown profile was used to allow multiple reactions with primer pairs of different annealing temperature to be run on the same thermal cycler. Thermal cycling was performed on a Tetrad PTC-225 Thermo Cycler (MJ Research).

Table 2.3 Generic PCR reaction master mix and volumes

Component	Volume per reaction (µl)
Megamix-W (Microzone)	22
Forward primer (20 nM; Sigma)	1
Reverse primer (20 nM; Sigma)	1
DNA (20 ng/µl)	1
TOTAL	25

Table 2.4 Thermal cycling conditions for generic PCR using a touch down profile

Step	Temperature	Time	Cycles
1 (Initial denaturation)	95 ^{o}C	5 minutes	1
2 (Denaturation)	95 ^{o}C	1 minute	10
3 (Primer annealing)	65 ^{o}C (-1 ^{o}C each cycle)	1 minute	
4 (Extension)	72 ^{o}C	1 minute	
5 (Denaturation)	95 ^{o}C	1 minute	20
6 (Primer annealing)	55 ^{o}C	1 minute	
7 (Extension)	72 ^{o}C	1 minute	
8 (Final extension)	72 ^{o}C	10 minutes	1

2.4 Agarose gel electrophoresis

Amplified PCR products were resolved on a 1.5% (w/v) agarose gel. The gel was made by dissolving 3 g of agarose (Sigma) in 200 ml of 1 x TBE buffer (89 mM Tris, 89 mM Boric Acid; 2 mM EDTA; pH8.3; Severn Biotechnology) by heating in a microwave at 400 W until fully dissolved. Once cooled 10 ng/ml ethidium bromide was added. Gels were cast in a 20 cm x 30 cm tray, 4 combs added and left to set. Gel electrophoresis was carried out in a horizontal tank containing 1 x TBE buffer. 5ul of PCR product was mixed with 5ul of 2x loading buffer (10 mg Orange G (Sigma), 1.5 ml Glycerol (Sigma) with ddH_2O added to a final volume of 25 ml) and run on the gel at 100 V for approximately 30 minutes. This allowed successfully amplified DNA products to be distinguished from those which had failed or where no specific amplification had occurred.

2.5 Sanger sequencing

The Sanger sequencing process was either performed manually by the author or by staff at the WMRGL using a robotically automated protocol. The methods used when the process was performed manually are described below.

2.5.1 PCR clean-up for Sanger sequencing

Successfully amplified PCR products underwent a PCR clean-up protocol during which a reaction was made up as per Table 2.5 and the enzymes activated and deactivated as per Table 2.6.

Table 2.5 PCR clean-up reaction master mix and volumes

Component	Volume per reaction (µl)
Antarctic Phosphatase Buffer (10 x; New England Biolabs)	1
Antarctic Phosphatase (New England Biolabs)	1
Exonuclease I (New England Biolabs)	0.25
PCR product	6
ddH_2O	1.75
TOTAL	10

Table 2.6 Thermal profile for PCR clean-up reaction

Step	Temperature	Time
1	37 ^oC	30 minutes
2	80 ^oC	20 minutes

2.5.2 Sequencing

Cleaned PCR reactions were Sanger sequenced. Each PCR product was sequenced in both the forward and reverse directions. The reaction master mix is shown in Table 2.7 and the thermal profile required in Table 2.8.

Table 2.7 Sequencing reaction master mix and volumes

Component	Volume per reaction (µl)
BigDye® buffer (5 x; Applied Biosystems)	2
Big Dye® v3.1 (Applied Biosystems)	0.5
M13 forward or reverse primer (20 µM; Sigma)	1
PCR product	4
ddH_2O	2.5
TOTAL	10

Table 2.8 Thermal profile for sequencing reaction

Step	Temperature	Time	Cycles
1 (Initial denaturation)	95 ^oC	5 minutes	1
2 (Denaturation)	96 ^oC	30 seconds	30
3 (Primer annealing)	50 ^oC	30 seconds	
4 (Extension)	60 ^oC	4 minutes	

2.5.3 Sequencing clean-up

An ethanol precipitation method was used to remove impurities from the reaction. The DNA was precipitated by addition of 1 µl of 250 mM EDTA (Sigma) and 35 µl of 100% ethanol (Sigma). Samples were centrifuged at 2500 rpm for 30 minutes at 4 ^{o}C to pellet the precipitated DNA, then ethanol removed by tapping samples upside down; the residual ethanol was removed by centrifuging plates upside down on filter paper at 400 rpm for 1 minute. The DNA pellet was washed with 200 µl of 70% (v/v) ethanol and centrifuged at 2500 rpm for 20 minutes at 4 ^{o}C to repellet the DNA. The ethanol was removed as previously and the pellet left to dry for 1 hour at room temperature. The DNA was mixed with 10 µl of HiDi™ formamide (Applied Biosystems), then denatured at 95 ^{o}C for 5 minutes, and then cooled on ice for 2 minutes.

2.5.4 Sequence data collection

Samples were then loaded onto an ABI 3730 capillary sequencer (Applied Biosystems).

2.5.5 Sequence data analysis

Data was analysed using Mutation Surveyor (SoftGenetics), against an appropriate gene specific reference file (GenBank file) downloaded from NCBI. Mutation Surveyor aids sequence analysis by comparing of the inputted electropherogram with the reference file and identifying discrepancies.

Chapter 3 Development of novel next generation sequencing (NGS) strategies for improved genetic diagnosis of inherited cancer predisposition diseases

Some of the work in this chapter has been reported in literature in the paper shown in section 9.1.

3.1 Introduction

As described in the introductory chapter, knowledge of cancer genetics has greatly advanced in the last 50 years. Many genes have been identified which cause predisposition to the development of a variety of tumours (2). Two examples of tumours that can be attributed to a genetic predisposition are phaeochromocytoma/paraganglioma and renal cell carcinoma (RCC; also known as kidney cancer). These tumour types are described below.

3.1.1 Phaeochromocytoma (PHEO) / Paraganglioma (PGL) / Head and neck paraganglioma (HNPGL)

A paraganglion is a group of cells derived from the neural crest. During differentiation paraganglia either become sympathetic and secrete catecholamines or parasympathetic and non-secretory. The term catecholamine describes a variety of naturally occurring amines that function as neurotransmitters and hormones within the human body. They are enzymatically synthesized from the amino acid l-tyrosine in the following sequence: tyrosine; dopa; dopamine; noradrenaline (norepinephrine); adrenaline (epinephrine). The majority of sympathetic paraganglia are located in the adrenal medulla, on top of the kidneys (43). Some sympathetic paraganglia also occur in the prevertebral and paravertebral sympathetic chains of the trunk and in the connective tissue within or near pelvic organs

(43). Parasympathetic paraganglia are almost always located in the head and neck, and located close to major arteries and nerves (43).

Paragangliomas are tumours of the paraganglia. Paragangliomas can originate in either parasympathetic or sympathetic paraganglia. The 'IARC WHO Classification of Tumours' defines a phaeochromocytoma (PHEO) as a functional (catecholamine secreting) paraganglioma derived from the adrenal medulla (44). However it should be noted that previously some authors used the term extra-adrenal phaeochromocytoma (pheochromocytoma) to describe a functional paraganglioma, henceforth referred to as PGL, arising outside the adrenal medulla (33). Sympathetic paraganglia-derived tumours (PHEO and PGL) almost always hypersecrete catecholamines, compared to only about 5% of parasympathetic PGLs (43). Those PGLs that do not secrete catecholamines are described as non-functional. Tumours derived from parasympathetic ganglia are referred to as head and neck paragangliomas (HNPGL) and, because they are overwhelmingly non-functional, their clinical presentation differs from that of PGL and PHEO.

Patients with PGL and PHEO (i.e. functional tumours) may present with headache, palpitations, sweating, nausea, flushing, weight loss, tiredness, hypertension, pallor, anxiety and/or panic attacks, with these symptoms often occurring paroxysmally (45). The majority of these symptoms are the result of the catecholamine production, with the paroxysmal nature of the symptoms related to episodic secretion of catecholamines by the tumour (45). By contrast, patients with non-functional tumours present with symptoms simply relating to the physical mass of the tumour compressing adjacent tissues and nerves (43).

PHEO/PGL/HNPGL have an estimated incidence of 1/300, 000 (43). The peak prevalence occurs at 40 years of age, with no evident gender bias (43). The majority of these tumours are benign but approximately 10% of PHEOs and 15-35% of PGLs are malignant (46). The majority of PHEOs present singly and if the tumour is going to metastasise this will normally occur within the first two years; however, metastasis can occur up to 40 years after the initial diagnosis (46). Patients with malignant PHEOs have a very poor prognosis, but it is very difficult to predict the malignant potential of a tumour when it is diagnosed (46). One study found a germline mutation in the *SDHB* gene to be present in 71.9% of paediatric patients with malignant tumours; in these patients the initial tumour was normally a functional PGL (47).

An inherited predisposition for PHEO/PGL/HNPGL can be caused by a pathogenic mutation in a number of different genes. At the commencement of this study, in 2012, a review of the genes associated with PHEO/PGL/HNPGL predisposition was performed (see section 3.4.1.1).

3.1.2 Renal cell carcinoma (RCC, also known as kidney cancer)

RCC was diagnosed in almost 340, 000 new patients in 2012, making it the 9th most common cancer worldwide (48). It mainly occurs in patients in their 60s and 70s, with double the number of males affected compared to females (49). There is a variable incidence of RCC throughout the world and the highest incidence is in developed countries (48). Over recent years there has been a rise in the frequency of RCC, this is thought to be only partially accounted for by improved screening and therefore partly to be the result of

an actual increase in disease (49). A small proportion (3-4%) of RCC is caused by an inherited susceptibility to disease, often in the context of a syndrome with other features (50). In comparison to sporadic cases, inherited RCC has an earlier age of diagnosis and is frequently bilateral and/or multicentric (49). There are a number of classes of RCC, as described in section 3.1.2.1. At the commencement of this study, in 2012, a review of the genes associated with RCC predisposition was performed (see section 3.4.1.2).

3.1.2.1 RCC classification

The main histopathological classes of RCC are listed below. Figure 3.1 provides examples of the typical histological findings.

3.1.2.1.1 Clear cell RCC (ccRCC)

ccRCC accounts for the vast majority of RCC cases, around 75% of tumours being of this type (49). When examined histologically, these tumours have clusters of cells with a clear cytoplasm, which results from the lipids in the cytoplasm dissolving during processing, and they are surrounded by a dense endothelial network (48).

3.1.2.1.2 Papillary RCC (pRCC)

pRCC makes up 10% of RCC cases (51). Histologically, these tumours have a papillary architecture, with basophilic cytoplasm and foamy histiocytes (48). There are two types of papillary tumour, referred to as Type I and Type II. Type I are small cell tumours, where the papillae are covered by small cells with little cytoplasm (51). Type II are higher grade with eosinophilic cytoplasm (51).

3.1.2.1.3 Chromophobe RCC (chRCC)

chRCC account for 5% of renal tumours and have the lowest risk of metastasis (51). Histologically, they have a large empty cytoplasm and perinuclear clearing (48). As compared to ccRCC the blood vessels are thick walled and eccentrically hyalinised (51).

3.1.2.1.4 Rare tumours

In addition, there are other, rare tumours that originate from the nephron and collecting duct. These are highly distinct from RCCs (48).

3.1.2.1.5 Oncocytoma (OC)

Oncocytomas have low malignant potential but are histologically similar to chromophobe RCC, which can complicate their diagnosis (48).

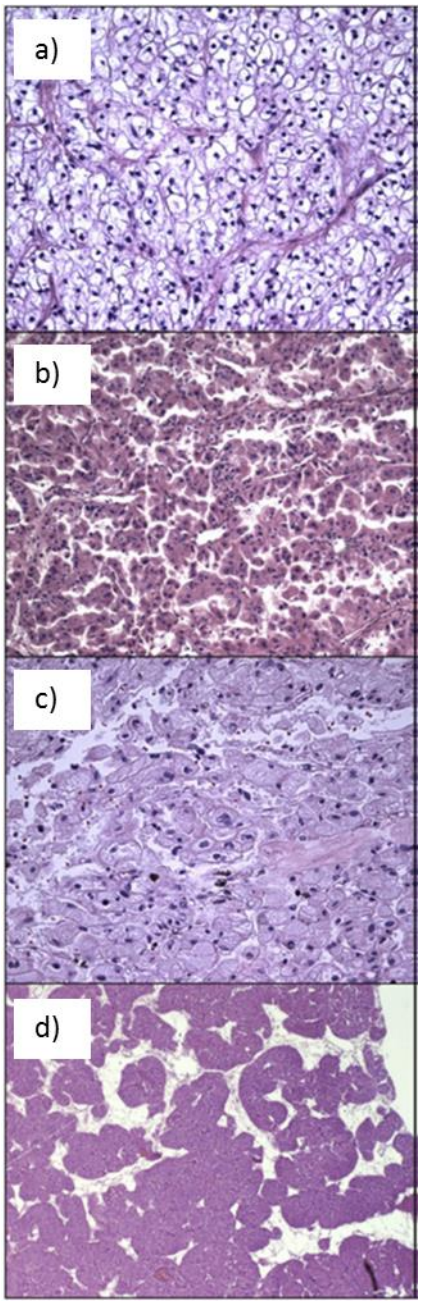


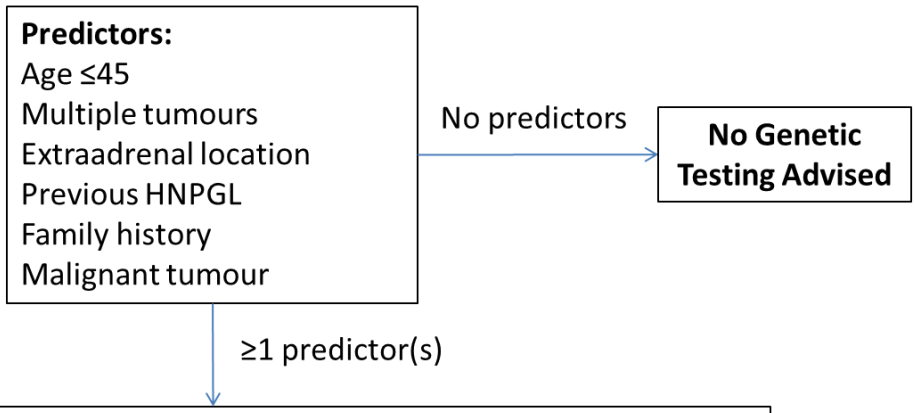
Figure 3.1 Examples of the typical histological findings in RCC: a) ccRCC; b) pRCC; c) chRCC; d) OC.
(Adapted from (51)).

3.1.3 Diagnostic genetic testing for PHEO/PGL/HNPGL and RCC predisposition

The importance of achieving a genetic diagnosis of an inherited cancer predisposition disease for both the proband and their extended family is discussed in the introductory chapter.

In 2012, when this study was conceived the vast majority of centres were offering testing of individual genes based on clinical request, with the lymphocyte DNA from an affected patient potentially undergoing multiple rounds of single gene analysis until the mutation was found or the cost became prohibitive. Clinicians used phenotype information combined with the prevalence of mutations within a gene to inform their strategy.

To take inherited PHEO/PGL/HNPGL as an example, it was proposed by a number of papers that all patients should be offered genetic testing despite only about a third of patients with a detectable mutation also having a positive family history (52;53). However, as the number of known causative genes increased, most centres adopted a targeted testing strategy to reduce costs (33;54;55). For example, it was recommended that testing for mutations in *RET*, *SDHB*, *SDHD* and *VHL*, was restricted to patients with one or more of the following: family history of PHEO/PGL/HNPGL, age <45 years, multiple tumours, extra-adrenal location and previous HNPGL (54). Algorithms were published for the order in which to test these genes to provide the lowest cost route to genetic diagnosis (see Figure 3.2) (54). Though such targeted testing strategies have a high mutation detection rate (% of tests that detect a mutation), a significant fraction of mutation positive individuals remain undetected (56).



Genetic Testing Advised			
Order	HNPGL	Single extra-adrenal tumour (not HNPGL)	All others
1.	*SDHD*	*SDHB*	*VHL*
2.	*SDHB*	*VHL*	*RET*
3.	*RET*	*SDHD*	*SDHB*
4.	*VHL*	*RET*	*SDHD*

Figure 3.2 Genetic testing algorithm for apparently non-syndromic PHEO/PGL/HNPGL cases
(Adapted from (54)).

The priority for genes to be analysed for patients with RCC was more easily determined as the tumour subtype combined with other phenotype information made gene selection a more straightforward process. However there were still individuals for whom the most likely causative gene was not the one containing the pathogenic mutation.

As discussed in Chapter One, the advent of massively parallel next generation sequencing (NGS) technologies has provided opportunities to radically alter strategies for diagnostic genetic testing. Compared to conventional Sanger sequencing, NGS provides increased capacity and speed at a significantly reduced cost (57). The use of molecular identifiers ("barcodes") allows multiplexing of samples, increasing throughput and efficiency (58;59).

Therefore the hypothesis explored in this chapter was that NGS could facilitate a transition from targeted, sequential analysis of individual PHEO/PGL/HNPGL or RCC genes in selected high risk individuals to a strategy of simultaneously testing multiple predisposition genes in all at-risk individuals. The technology available to test this hypothesis was the Fluidigm Access Array (AA) and the Roche GS Junior.

3.2 Aims

The aims of the research reported in this chapter of the thesis were to:

1. Perform a literature review to identify the genes known to cause inherited PHEO/PGL/HNPGL and RCC;
2. Design and optimise the strategy selected to analyse the identified genes;
3. Validate the novel PHEO/PGL/HNPGL gene panel testing strategy;
4. Validate the novel RCC gene panel testing strategy;
5. Use the novel NGS assays for prospective patient testing.

3.3 Materials and methods

3.3.1 Literature review

A review of the literature was performed in January 2012 to identify all genes known to cause a) PHEO/PGL/HNPGL and b) RCC. A combination of the word 'gene' plus 1) 'paraganglioma', 2) 'phaeochromocytoma', 3) 'pheochromocytoma', 4) 'renal cell cancer', 5) 'renal cell carcinoma' and 6) 'kidney cancer' were entered in the NCBI PubMed database search engine. All returns were examined for novel genes reported to be causing inherited disease and each gene reviewed. In addition two recent review articles were examined:

a) Jafri M, Maher ER (2012) The genetics of phaeochromocytoma: using clinical features to guide genetic testing. Eur. J. Endocrinol. 166(2):151-8 (33);

b) Verine J, Pluvinage A, Bousquet G, Lehmann-Che J, de Bazelaire C, Soufir N, Mongiat-Artus P. (2010) Hereditary renal cancer syndromes: an update of a systematic review. Eur. Urol. 58(5):701-10 (60).

3.3.2 Subjects

This study was formed of three main phases: 1) optimisation; 2) validation and 3) prospective analysis.

In the first and second phases, DNA samples which contained variants previously identified by Sanger sequencing were studied. For the PHEO/PGL/HNPGL panel, 85 DNA samples containing known variant calls were run. Of the 182 variants which were analysed, 84 were unique. They comprised 59 single base substitutions, 14 deletions, 6 duplications, 3 insertions and 2 indels. For the RCC panel, 42 DNA samples containing known variant calls

were run. There were 106 variants analysed of which 45 were unique. The unique variants comprised 31 substitutions, 9 deletions, 4 duplications and 1 insertion.

In the third phase, 239 DNA samples were analysed prospectively. For the PHEO/PGL/HNPGL panel, 120 DNA samples for patients referred with PHEO/PGL/HNPGL were prospectively tested. Of these patients, 42 (35%) had previously had 1 or more genes on the panel tested by Sanger sequencing and either no variants identified or a variant of uncertain significance (VUS). The remainder had not undergone any previous genetic analysis. For the RCC panel, 119 DNA samples from patients with RCC were tested prospectively. Of these patients, 19% had had one or more genes analysed previously by Sanger sequencing and analysis of the additional genes on the panel was requested. These patients were referred from genetic centres around the United Kingdom.

3.3.3 Assay design

The technology to be used during this phase of the research was determined by what was locally available. For amplicon-based target enrichment, the 48.48 Access Array (AA) system (Fluidigm Corporation) was used. The available NGS sequencer was the GS Junior (Roche 454 Life Sciences), which is a 454 pyrosequencing-based technology (61).

3.3.3.1 PHEO/PGL/HNPGL panel

The assay included 9 genes identified to be associated with inherited predisposition to PHEO/PGL/HNPGL following the literature review (section 3.4.1):

- *MAX* (NM_002382.3),

- *RET* (NM_020975.4),
- *SDHA* (NM_004168.2),
- *SDHAF2* (NM_017841.2),
- *SDHB* (NM_003000.2),
- *SDHC* (NM_003001.3),
- *SDHD* (NM_003002.2),
- *TMEM127* (NM_017849.3) and
- *VHL* (NM_000551.3).

Primer pairs were designed using Primer3 (62) to amplify the full coding sequence of all genes, except for the RET gene for which only exons 10, 11 and 13-16 were included. Amplicons were designed following Fluidigm AA recommendations, amplicon sizes were within a 150 bp range (310-460 bp), melting temperatures were within a 2 ^{o}C range (59-61 ^{o}C) and primers had no more than 3 of the same base in a run, whenever possible. To avoid amplification of *SDHA* pseudogenes, primers were designed so that the 3' nucleotide of at least one but preferably both primers was specific to *SDHA*. The primer sequences are provided in Appendix Table 7.1.

3.3.3.2 RCC panel

The assay included 5 genes identified to be associated with inherited predisposition to RCC following the literature review (section 3.4.1):

- *FLCN* (NM_144997.5),
- *FH* (NM_000143.3),
- *MET* (NM_000245.2),

- *SDHB* (NM_003000.2) and
- *VHL* (NM_000551.3).

Primer pairs were designed using Primer3 (62), again as per the Fluidigm specifications. Amplicon sizes ranged from 338-451 bp, had a melting temperature of 59-61 ^{o}C and primers had no more than 3 of the same base in a run, whenever possible. The primer sequences are provided in Appendix Table 7.1.

3.3.4 Target enrichment and multiplexed next generation sequencing (NGS)

Genomic DNA was amplified using the 48.48 AA system, which facilitates the simultaneous amplification of at least 48 targets for up to 48 patient DNA samples. Patient specific barcodes and the 454 sequencing adaptors were added during a second round of PCR. All processes were performed according to the manufacturer's protocol 'Multiplex Amplicon Tagging for 454 Titanium Sequencing on the 48.48 Access Array IFC'. The patient samples were then pooled in equal volumes prior to NGS on the GS Junior. Emulsion PCR, bead enrichment and pyrosequencing were carried out as per the manufacturer's instructions. For the PHEO/PGL/HNPGL panel, the Lib-A kit with 0.5:1 molecule to bead ratio was used following optimisation and for the RCC panel the Lib-A kit with 1:1 molecule to bead ratio was used following optimisation.

All optimisation and validation runs and some diagnostic runs were performed by the author. Some diagnostic runs were performed by Anna Yeung and Lindsey Vialard on behalf of the WMRGL.

3.3.5 NGS data analysis

Data analysis was performed using NextGENe v2.16 sequence analysis software (SoftGenetics, State College, PA, USA) and a semi-automated in-house bioinformatics pipeline. Poor quality reads were removed, samples were demultiplexed and reads mapped against gene-specific GenBank reference files. Variants were then filtered step-wise, so variants were only considered when:

- Within the coding sequence ± 5 base pairs;
- In >15% of >15 total reads for the base being interrogated. Regions below 30-fold coverage were analysed by Sanger sequencing. The lower filtering parameter allow identification of likely variants;
- Not a known benign polymorphism;
- A defined threshold was exceeded (for insertion and deletion variants only):
    - Analysis of indels was performed using NextGENe variant comparison tool (VCT) and the in house pipeline, as a result of the issues surrounding homopolymer tract calls with the 454 chemistry. An indel variant was only included when it was not within 2 standard deviations of the mean percentage call for that variant for all patients in that run.

3.3.6 Sanger sequencing

PCR of exons requiring further analysis was followed by bidirectional Sanger sequencing using the ABI BigDye® v3.1 Terminator Cycle Kit (Applied Biosystems) and then analysed

using an ABI Prism 3730 (Applied Biosystems) and the Mutation Surveyor (SoftGenetics) program. See Chapter 2 for additional details.

3.3.7 Multiplex ligation-dependent probe amplification

The MLPA probemixes SALSA P256 FLCN, SALSA P226 SDH and SALSA P016 VHL (MRC Holland) were used for dosage analysis, as per the manufacturer's instructions, and the data analysed using GeneMarker v1.70 (SoftGenetics) program.

3.3.8 Variant assessment

Variant pathogenicity assessment was performed as per best practice guidelines (63). The majority of the assessment used Alamut as an interface (Interactive Biosoftware) and involved a) interrogation of locus specific databases, the Human Gene Mutation Database and in house databases, b) literature review, c) search of dbSNP, NHLBI Exome Sequencing Project Exome Variant Server and 1000 genomes project databases, d) species conservation and e) *in silico* prediction for the impact of missense and splicing variants, as appropriate.

3.4 Results

3.4.1 Literature review

3.4.1.1 Genes known to cause an inherited predisposition to PHEO/PGL/HNPGL (January 2012)

In January 2012, there were 12 genes known to be associated with PHEO/PGL/HNPGL. This section provides basic details of these genes in terms of PHEO/PGL/HNPGL susceptibility.

3.4.1.1.1 The *EGLN1* gene

Germline mutations in *EGLN1* have been associated with congenital erthyrocytosis (64). In 2008, a single case of a patient with erthyrocytosis and a recurrent abdominal PGL was reported (65). Upon sequencing *EGLN1* the patient was found to have a sequence variant; c.1121A>G (p.His374Arg), which was postulated to be causative as it showed loss of heterozygosity (LOH) in tumour studies, and functional assays showed it to affect EGLN1 and stabilise HIF-α proteins (65). However, there have been no subsequent reports of mutations in *EGLN1* being associated with PHEO/PGL/HNPGL. Astuti *et. al.* carried out sequencing of the three 2-oxoglutarate-dependent prolylhydroxylase enzymes, including *EGLN1*, in 82 patients with inherited PHEO/PGL/HNPGL and did not find any mutations (64). This suggests that mutations in these genes are not a common cause of inherited PHEO/PGL/HNPGL.

3.4.1.1.2 The *KIF1B* gene

In 2008, Schlisio *et. al.* reported missense mutations in *KIF1B* in 2 patients with PHEO (66). However, since then no additional reports of mutations in this gene have been published. It is therefore unclear if this is a true PHEO susceptibility gene. Zhao *et. al.* reported a family

with a loss of function missense mutation in *KIF1B* causing Charcot-Marie-Tooth type 2A1 (67).

3.4.1.1.3 The *MAX* gene

In 2011, Comino-Méndez *et. al.* identified 3 pathogenic mutations in *MAX* during exome sequencing of individuals with inherited PHEO (68). The same study then looked for mutations in *MAX* in 59 additional patients and identified another 5 mutations (68). Analysis of mutation positive tumours identified acquired isodisomy for the *MAX* gene (68). *MAX* is part of the MYC-MAX-MXD1 network of transcription factors (68).

3.4.1.1.4 The *NF1* gene

The *NF1* gene is a tumour suppressor gene (TSG). Its most well characterised function is as a GTPase that inactivates RAS which in turn inhibits the MAPK signalling cascade (69). Heterozygous mutations in the *NF1* gene cause autosomal dominant neurofibromatosis type 1 (NF1), which is a characterised by pigmentary abnormalities and the neoplastic growth of neural crest-derived cells (70). PHEOs are observed in 0.1-5.7% of patients with NF1, always in conjunction with other symptoms (70). Bauch *et. al.* could not establish a correlation between a specific *NF1* genotype and PHEO, although they did note that patients with PHEO were more likely to have a mutation in the cysteine rich domain of the protein (70).

3.4.1.1.5 The *RET* gene

The *RET* gene is a proto-oncogene which acts as a receptor tyrosine kinase. Ligand binding induces dimerisation of the RET protein which leads to activation of multiple signal transduction cascades (71). The *RET* gene shows phenotypic heterogeneity, causing two clinically distinct diseases, multiple endocrine neoplasia type 2 (MEN2)/familial medullary

thyroid cancer (FMTC) and Hirschsprung disease. In Hirschsprung disease there is congenital absence of the enteric neurons in the gastrointestinal tract, whereas MEN2 comprises diseases of the endocrine system. MEN2/FMTC can be subdivided in to 3 types (see Table 3.1). PHEO is observed in two of those types, and is often the presenting symptom. Hirschsprung disease is caused by inactivating mutations which are spread throughout the gene, whereas MEN2/FMTC is caused by activating mutations, which have only been observed in exons 10, 11, 13, 14 and 16 (listed in Table 3.1). Therefore, mutation screening can be limited to specific codons or exons for MEN2 (21).

Table 3.1 Classes of MEN2
(Adapted from (21)).

Type of MEN2	Symptoms	Associated mutations
Familial medullary thyroid cancer (FMTC)	medullary thyroid cancer only - onset in middle age	R321G, G533C, R600Q, K603E, Y606C, C609R/G/S/Y, C611R/G/F/S/W/Y, C618R/G/F/S/Y, C620R/G/F/S/W/Y, C630R/F/S/Y, C634R/G/F/S/W/Y, E768D, N777S, V778I, Q781R, L790F, Y791F, V804L/M, I852M, S891A, R912P
MEN2A	PHEO; Parathyroid adenoma or hyperplasia; Medullary thyroid cancer - onset in early adulthood	C609R/G/S/Y, C611R/G/F/S/W/Y, C618R/G/F/S/Y, C620R/G/F/S/W/Y, C630R/F/S/Y, C634R/G/F/S/W/Y, K666E, E768D, L790F, Y791F, V804L/M, S891A
MEN2B	PHEO; Parathyroid adenoma or hyperplasia; Medullary thyroid cancer - onset in early childhood; Mucosal neuromas of the lips and tongue; Distinctive facies with enlarged lips; ganglioneuromatosis of the gastrointestinal tract; Asthenic "Marfanoid" body habitus.	A883F, M918T

3.4.1.1.6 The *SDHA* gene

The *SDHA* gene encodes one of 4 subunits of succinate dehydrogenase (SDH), a mitochondrial enzyme that participates in both the citric acid cycle and the electron transport chain. Biallelic *SDHA* mutations have been shown to cause an early onset encephalopathy called Leigh syndrome but the gene was not thought to have a role in PHEO/PGL/HNPGL formation (72). Burnichon *et. al.* identified a likely pathogenic missense mutation in a patient with a PGL in 2010 (72). Subsequently, the same group published a study in which they identified an additional 5 patients with a germline mutation in *SDHA* (73). *SDHA* analysis is complicated by its four known pseudogenes, generated by complete or partial gene duplications (74).

3.4.1.1.7 The *SDHB* gene

SDHB is the second of the four subunits of SDH. In 2001, it was shown that heterozygous loss of function mutations in *SDHB* could predispose an individual to PHEO/PGL/HNPGL (30). *SDHB* mutation carriers are at higher risk of developing abdominal PHEO/PGL than HNPGL (75). In addition *SDHB* mutation carriers have a higher risk of malignant disease (75).

3.4.1.1.8 The *SDHC* gene

SDHC encodes the third subunit of SDH. Mutations in *SDHC* were first shown to cause inherited PGL in 2000 (76). Mutations in *SDHC* are comparatively rare, compared to *SDHB* and *SDHD*. A 2004 study found that they only predispose to HNPGL and not PHEO/PGL (77). However there has since been a report of a patient with an *SDHC* mutation who has PGL (78).

3.4.1.1.9 The *SDHD* gene

SDHD is the fourth of the SDH subunits but was the first of the SDH subunit genes to be associated with PHEO/PGL/HNPGL (79). Mutations in *SDHD* are more likely to predispose to non-functional HNPGL than functional PHEO/PGL (75). Inheritance of disease differs from the conventional autosomal dominant manner expected for mutations in cancer predisposition genes. Although more than one case of maternal transmission with disease in offspring is referenced in the literature (80), it is rare for maternal transmission of a pathogenic *SDHD* mutation to cause disease in offspring (80). Therefore the *SDHD* gene generally follows a mode of inheritance consistent with genomic imprinting (80), however, it sits in a chromosomal region that has been shown to have biallelic expression in several tissues (80).

3.4.1.1.10 The *SDHAF2* gene

The *SDHAF2* gene encodes SDH complex assembly factor 2, a co-factor of flavin adenine dinucleotide (FAD) insertion into the SDH complex (81). *SDHAF2* has a role in flavination of SDHA, which is essential for a fully functional SDH complex (81). To date, three families have been identified with mutations and, like *SDHD*, *SDHAF2* appears to show parent of origin specific expression with disease only being expressed when the mutation is inherited from the father (81-83). Both *SDHD* and *SDHAF2* are located on the short arm of chromosome 11.

3.4.1.1.11 The *TMEM127* gene

TMEM127 is a transmembrane protein which is part of the mTORC1 signalling complex, a critical control junction for protein synthesis and cell survival (84). A large multicentre study found that germline mutations of *TMEM127* were responsible for disease in around 2% of

the patients examined, but were associated only with PHEOs and not PGL or HNPGLs (84). However, subsequent investigation has found likely pathogenic *TMEM127* mutations in one patient with a HNPGL and one patient with a retroperitoneal PGL (85).

3.4.1.1.12 The *VHL* gene

The *VHL* gene is a TSG. Mutations in the *VHL* gene cause von Hippel-Lindau (VHL) disease. VHL disease can be categorised on the basis of the tumour types observed (see Table 3.2). PHEO is the distinguishing factor between Type 1 and Type 2 disease, with PHEO being the only manifestation in Type 2C. PGL is observed infrequently (60). For additional information on VHL refer to Chapter 4, where a detailed commentary is provided.

Table 3.2 The types of VHL
HAB = retinal and CNS haemangioblastomas, RCC = renal cell carcinoma, PHEO = pheochromocytoma (Adapted from (86)).

VHL disease type	Tumour types observed		
	HAB	RCC	PHEO
Type 1	+	+	-
Type 2A	+	-	+
Type 2B	+	+	+
Type 2C	-	-	+

3.4.1.1.13 Conclusion

The assay was intended to identify patients with either apparently sporadic PHEO/PGL/HNPGL or familial PHEO/PGL/HNPGL, and therefore any genes which only cause PHEO/PGL/HNPGL as part of a clear syndrome or where PHEO/PGL/HNPGL is never the presenting symptom were not included. The *NF1* gene was ruled out of the assay as PHEO are only seen in a small percentage of cases of NF1 and the clinical features of NF1 are invariably already in place by the time PHEO presents (mean age 40 years) (53).

The remaining genes were prioritised according to potential for mutation identification. Table 3.3 lists the genes that have been shown to cause inherited and/or syndromic PHEO and/or PGL and/or HNPGL and gives the reported frequency of mutations identified in each gene in patients with an apparently sporadic PHEO/PGL/HNPGL. *EGLN1* and *KIF1B* did not have frequency data available and the initial report was several years old with no subsequent publications, suggesting that the identification of pathogenic mutations had not been replicated in other centres. Therefore these genes were not included.

Table 3.3 Frequency of mutations identified in apparently sporadic PHEO/PGL
This table does not included familial mutation frequency.

Gene	Exons	Frequency of mutations identified in apparently sporadic PHEO/PGL/HNPGL (Reference)
EGLN1	5	No data
KIF1B	47	No data
MAX	5	1.2% (28)
RET	5*	0.4%-5.0% (87)
SDHA	15	3% (73)
SDHAF2	4	0% (81)
SDHB	8	1.5%-10.0% (87)
SDHC	6	0% (77)
SDHD	4	0.8%-10.0% (87)
TMEM127	4	2% (84)
VHL	3	3.5%-11.1% (87)

* 5/20 exons have been shown to harbour mutations associated with MEN2

3.4.1.2 Genes known to cause an inherited predisposition to RCC (January 2012)

The sections below briefly describe the genes known to cause inherited predisposition to renal cell carcinoma, as of January 2012.

3.4.1.2.1 The *VHL* gene

VHL is the most common hereditary renal cancer syndrome (60). RCC is found in Type 1 and Type 2B VHL disease and in 24–45% of VHL patients (60). Almost all the tumours are ccRCCs (88). For additional information on VHL refer to Chapter 4, where a detailed commentary is provided.

3.4.1.2.2 The *MET* gene

The *MET* gene is a proto-oncogene situated on chromosome 7q31.3. It encodes the cell surface receptor for Hepatocyte Growth Factor (HGF). Predisposition to hereditary papillary renal carcinoma (HPRC) is caused by activating mutations in the tyrosine kinase domain of *MET* (89).

3.4.1.2.3 The *FH* gene

The *FH* gene encodes the fumarate hydratase protein, an enzyme that catalyses the conversion of fumarate to malate in the Krebs cycle (90). Mutations in this gene cause hereditary leiomyomatosis and renal cell cancer (HLRCC) (90). RCC is found in approximately 20% of HLRCC families (60). pRCC or collecting duct tumours are typically observed and tend to be aggressive and thus require immediate removal upon detection (49).

3.4.1.2.4 The *FLCN* gene

FLCN is a TSG involved in the regulation of adenosine monophosphate (AMP)-activated protein kinase (AMPK) and mTOR signalling pathways (91). Mutations in the *FLCN* gene cause Birt-Hogg-Dubé (BHD) syndrome, which is characterised by the development of benign skin tumours, fibrofolliculomas, on the face and upper torso, along with increased susceptibility to spontaneous pneumothorax and RCC (92). Renal tumours occur in 25–35% of BHD patients (60). BHD-related renal tumours can have different histologic subtypes, chRCCs and hybrid oncocytic tumours (mixed pattern of chRCCs and OCs) are most frequently observed but other types, including clear cell, are also seen (60).

3.4.1.2.5 The *TSC1* and *TSC2* genes

Tuberous sclerosis (TS) is caused by an inactivating mutation in *TSC1* or *TSC2* (93;94). *TSC1* encodes the protein hamartin and *TSC2* encodes tuberin. Hamartin and tuberin interact to produce a heterodimer and together function as part of the mTOR pathway. TS is a multisystem disorder characterised by the formation of hamartomas in multiple organ systems, such as the heart, brain, skin, lungs and kidneys. Renal lesions occur in 50–80% of TS patients (60). However the majority of these lesions are angiomyolipomas with RCC only reported in 1–4% (60). This is similar to the general population rate of RCC but in TS patients the RCC occurs at a younger age (60). The RCCs observed are predominately ccRCCs, but other RCC types have also been recorded (chRCCs, pRCCs, and OCs). RCC is not seen alone in TS (60).

3.4.1.2.6 The *CDC73* gene

CDC73 is a TSG which encodes a protein called parafibromin. One of parafibromin's roles is to act as part of the PAF multi-protein complex which associates with the large subunit of RNA polymerase II and a histone methyltransferase complex (95). Mutations in *CDC73* cause hereditary hyperparathyroidism-jaw tumour (HPT-JT) syndrome which is characterised by a predisposition to develop primary hyperparathyroidism and multiple ossifying jaw fibroma (96). A variety of renal manifestations have been described. They occur in around 15% of patients and include RCCs (60). RCC is not the predominant or presenting feature.

3.4.1.2.7 The *SDHB* gene

Mutations in SDHB are typically implicated in PHEO/PGL/HNPGL predisposition (see section 3.4.1.1.7); however, renal tumours, including ccRCCs, chRCCs and OCs have also been identified in up to 5% of patients and can be the presenting feature (75).

3.4.1.2.8 Conclusions

Of the eight genes discussed above, only five have been shown to give rise to RCC as the presenting feature of disease. The *TSC1/TSC2* and *CDC73* genes were not included in NGS panel design because the RCC should always be recognised in the context of the wider clinical picture. Therefore *VHL, FH, FLCN, MET* and *SDHB* were chosen to be included in the assay.

3.4.2 Assay design and optimisation

3.4.2.1 The PHEO/PGL/HNPGL panel

Assay design and optimisation was performed initially for the PHEO/PGL/HNPGL panel. A total of 55 primer pairs for the Access Array (AA) were designed for the selected genes and trialled using scaled up volumes of the reactions. All primer pairs had to amplify DNA at common reaction conditions, as the PCR occurred within the AA on a single thermal cycler. This meant that primer redesign was the only option for poorly performing primer pairs. Six amplicons were redesigned to optimize amplification as a result of non-specific amplification or excessive primer dimer. The majority of primers were then successful but despite redesign, RET exon 14 did not produce an amplicon of sufficient quality and it was removed from the AA set-up to be analysed by Sanger sequencing.

Early NGS runs tested assay design and sequencing depth per fragment, and optimised the number of patients per run to give >30-fold coverage for all amplicons. Four amplicons producing consistently low reads were removed from the assay to be analysed by Sanger sequencing. These amplicons had >66% GC content: *MAX* exon 1 (67%); *SDHA* exon 1 (75%); *SDHAF2* exon 1 (68%) and *TMEM127* exon 2 (72%). This left 49 amplicons for the AA.

The AA was adjusted to allow the most uniform NGS coverage possible. Six amplicons with consistently high read numbers were diplexed within the microreactions on the 48.48 AA, with diplexed amplicons matched for size and melting temperature as closely as reasonably possible. Those with low read numbers were amplified in duplicate (n=2). It was established that higher sequencing depth was achieved when a 2 stage PCR was employed (stage 1 target specific PCR, stage 2 indexing PCR) rather than the initially trialled single reaction

'step-out' PCR. The optimal number of patients sequenced per run in order to attain >30-fold coverage for all amplicons was determined to be 20 and the optimal DNA molecule to sequencing bead ratio for the NGS was determined to be 0.5:1.

During the optimization and validation phase a mean number of 82, 000 passed filter reads was achieved for the PHEO/PGL/HNPGL panel.

3.4.2.2 The RCC panel

Lessons from the design and optimisation of the PHEO/PGL/HNPGL panel were applied to the RCC panel. Amplicons with greater than 66% GC content were not included on the AA assay and the 2-stage PCR for target amplification and indexing was applied from the outset. *FH* exon 9 did not work as part of this system and it instead was analysed by Sanger sequencing. The reason for this was unclear, as it did work at a scaled up reaction volume, however it did not produce any mapped reads on the validation runs. The optimal number of patients sequenced per run in order to attain >30-fold coverage for all amplicons was determined to be 24, as a result of the smaller number of amplicons. Finally, the optimal DNA molecule to sequencing bead ratio for the NGS was determined to be 1:1.

3.4.2.3 PHEO/PGL/HNPGL panel validation

3.4.2.3.1 PHEO/PGL/HNPGL panel sensitivity

The NGS assay was validated using 85 patient DNAs containing 182 variants (84 unique variants) (Table 3.4). 171 variants met the required minimum coverage threshold set (>30x depth) and of these 170 were detected (76/77 unique variants), giving an overall sensitivity of 98.7% (95% CI: 92.95% to 99.78%).

The single undetected variant (with >30-fold coverage) was a 6 bp duplication in *SDHB* intron 4, c.424-19_424-14dupTTCTTC. This is a polymorphic tract where the major allele contains 8 TTC repeats. The duplication allele was correctly identified in 7/49 (14.3%) reads. A number of sequencing/alignment errors had occurred leading to the incorrect calling. This polymorphism was sequenced on an additional 5 occasions and detected each time (mean = 26% of reads).

Table 3.4 Unique variants used to validate the PHEO/PGL/HNPGL panel and MiSeq technology

Gene	Exon/ Intron	Nucleotide change	Protein change	Variant type
RET	9	c.1760-158C>G		Substitution
RET	10	c.1853G>C	p.Cys618Ser	Substitution
RET	11	c.1900T>C	p.Cys634Arg	Substitution
RET	11	c.2071G>A	p.Gly691Ser	Substitution
RET	13	c.2307T>G	p.=	Substitution
RET	14	c.2410G>A	p.Val804Met	Substitution
RET	14	c.2608-24G>A		Substitution
RET	15	c.2712C>G	p.=	Substitution
RET	16	c.2801+54A>T		Substitution
SDHB	1	c.17_42dup26	p.Ala15Hisfs*4	Insertion/ Deletion
SDHB	1	c.18A>C	p.=	Substitution
SDHB	1	c.72+1G>A		Substitution
SDHB	1	c.72+1G>T		Substitution
SDHB	1	c.79C>T	p.Arg27*	Substitution
SDHB	2	c.136C>T	p.Arg46*	Substitution
SDHB	2	c.137G>A	p.Arg46Gln	Substitution
SDHB	2	c.166_170delCCTCA	p.Pro56Tyrfs*5	Insertion/ Deletion
SDHB	2	c.200+33G>A		Substitution
SDHB	2	c.200+35G>A		Substitution
SDHB	2	c.200+104G>A		Substitution
SDHB	2	c.88delC	p.Gln30Argfs*47	Insertion/ Deletion
SDHB	2	c.201-36T>G		Substitution
SDHB	3	c.282_283insCTTA	p.Glu95Leufs*25	Insertion/ Deletion
SDHB	3	c.286+169A>G		Substitution
SDHB	3	c.286G>A	p.Gly96Ser	Substitution
SDHB	3	c.287-1G>C		Substitution
SDHB	4	c.297delC	p.Ser100Leufs*4	Insertion/ Deletion
SDHB	4	c.311delAinsGG	p.Asn104Argfs*15	Insertion/ Deletion
SDHB	4	c.338G>A	p.Cys113Tyr	Substitution
SDHB	4	c.379dupA	p.Ile127Asnfs*28	Insertion/ Deletion
SDHB	4	c.424-19_424-14dup6		Insertion/ Deletion

Gene	Exon/ Intron	Nucleotide change	Protein change	Variant type
SDHB	5	c.502dupC	p.Gln168Profs*11	Insertion/ Deletion
SDHB	6	c.470T>A	p.Val157Asp	Substitution
SDHB	6	c.553G>T	p.Glu185*	Substitution
SDHB	6	c.587G>A	p.Cys196Tyr	Substitution
SDHB	6	c.591delC	p.Ser198Alafs*22	Insertion/ Deletion
SDHB	7	c.685_686ins13	p.Glu229Alafs*31	Insertion/ Deletion
SDHB	7	c.689G>A	p.Arg230His	Substitution
SDHB	7	c.745_748dupTGCA	p.Thr250Metfs*7	Insertion/ Deletion
SDHB	8	c.801G>T	p.Lys267Asn	Substitution
SDHC	1	c.20+11_20+12dupTG		Insertion/ Deletion
SDHC	2	c.21-189A>G		Substitution
SDHC	2	c.21-211C>T		Substitution
SDHC	2	c.21-96C>T		Substitution
SDHC	3	c.148C>T	p.Arg50Cys	Substitution
SDHC	4	c.214C>T	p.Arg72Cys	Substitution
SDHC	4	c.241+74C>T		Substitution
SDHC	4	c.242-193A>G		Substitution
SDHC	4	c.242-174T>C		Substitution
SDHC	4	c.242-154A>G		Substitution
SDHC	4	c.242-136G>T		Substitution
SDHC	5	c.397C>T	p.Arg133*	Substitution
SDHD	1	c.14G>A	p.Trp5*	Substitution
SDHD	1	c.47_51+1del7		Insertion/ Deletion
SDHD	1	c.52+136G>T		Substitution
SDHD	2	c.144_145dupCA	p.Ile49Thrfs*38	Insertion/ Deletion
SDHD	2	c.57delG	p.Leu20Cysfs*66	Insertion/ Deletion
SDHD	2	c.94_95delTC	p.Ala33Ilefs*35	Insertion/ Deletion
SDHD	2	c.170-29A>G		Substitution
SDHD	3	c.204C>T	p.=	Substitution
SDHD	3	c.205G>T	p.Glu69*	Substitution
SDHD	3	c.242C>T	p.Pro81Leu	Substitution
SDHD	3	c.274G>T	p.Asn92Tyr	Substitution

Gene	Exon/ Intron	Nucleotide change	Protein change	Variant type
SDHD	3	c.276_278delCTA	p.Tyr93del	Insertion/ Deletion
SDHD	3	c.296delT	p.Leu99Profs*36	Insertion/ Deletion
SDHD	3	c.315-32T>C		Substitution
SDHD	4	c.342T>A	p.Tyr114*	Substitution
VHL	1	c.214delT	p.Ser72Profs*87	Insertion/ Deletion
VHL	1	c.239G>T	p.Ser80Ile	Substitution
VHL	1	c.326_340+16del31insA		Insertion/ Deletion
VHL	1	c.337delC		Insertion/ Deletion
VHL	1	c.340+5G>C		Substitution
VHL	2	c.293A>C	p.Tyr98Ser	Substitution
VHL	2	c.358A>T	p.Arg120*	Substitution
VHL	2	c.376G>A	p.Asp126Asn	Substitution
VHL	2	c.381delG,	p.Leu128Phefs*31	Insertion/ Deletion
VHL	2	c.463+108T>G		Substitution
VHL	2	c.463+174delT		Insertion/ Deletion
VHL	3	c.479_480delAG	p.Glu160Alafs*13	Insertion/ Deletion
VHL	3	c.482G>C	p.Arg161Pro	Substitution
VHL	3	c.499C>T	p.Arg167Trp	Substitution
VHL	3	c.548C>T	p.Ser183Leu	Substitution
VHL	3	c.562C>G	p.Leu188Val	Substitution
VHL	3	c.598C>T	p.Arg200Trp	Substitution

3.4.2.3.2 PHEO/PGL/HNPGL panel specificity

False variant calls resulting from errors in determining the number of bases within homopolymer tracts is a recognized problem with the 454 sequencing chemistry (97). Inter-run variability was observed for homopolymer errors, with the recurrent errors observed within a run, varying between runs. During the validation period, numerous false positive calls were identified for every DNA sample. In one sequencing run there were 164 unique

variants and each DNA sample had an average of 65 variants called, an average of 46 of which were likely homopolymer related artefacts. To limit the Sanger sequencing confirmations required and to increase specificity, two filtering steps were introduced: 1. the region of interest was limited to the coding sequence ±5 bp (with the exception of 5' and 3' untranslated regions). This reduced the number of unique variants identified to 27 and only 4 were likely artefacts associated with homopolymer tracts. These 4 variant calls comprised 36% of all variants observed. 2. The in-house bioinformatics pipeline was set to filter out homopolymer related deletions/duplications that were within two standard deviations of the mean for that variant for the patients on that run. It also filtered benign polymorphisms leaving an average of 0.5 variants per DNA sample which required Sanger confirmation and pathogenicity analysis. Performing this filtering on the validation runs did not cause any known pathogenic mutations to be missed.

3.4.2.3.3 PHEO/PGL/HNPGL variant frequency

Variant frequency is calculated by dividing the number of mutant reads by the total number of reads. Figure 3.3 provides details of the variant frequency observed for the 171 variants input into the system. The average variant frequency was close to that expected for all variant types. The heterozygous substitution variants were seen in an average of 49.6% of reads (range: 16.2-94.0%) and the heterozygous insertion/deletion variants in an average of 44.10% (range: 14.3-73.0%). The homozygous substitutions were seen at an average of 99.4% (range: 89.3-100.0%). Outliers comprised the 6 base pair duplication discussed above and a heterozygous substitution observed at 94.0%, which upon investigation was considered likely to be the result of the sample being degraded. Also, 5/26 homozygous

substitutions were not identified in 100% of reads. This was caused by low level sequencing errors.

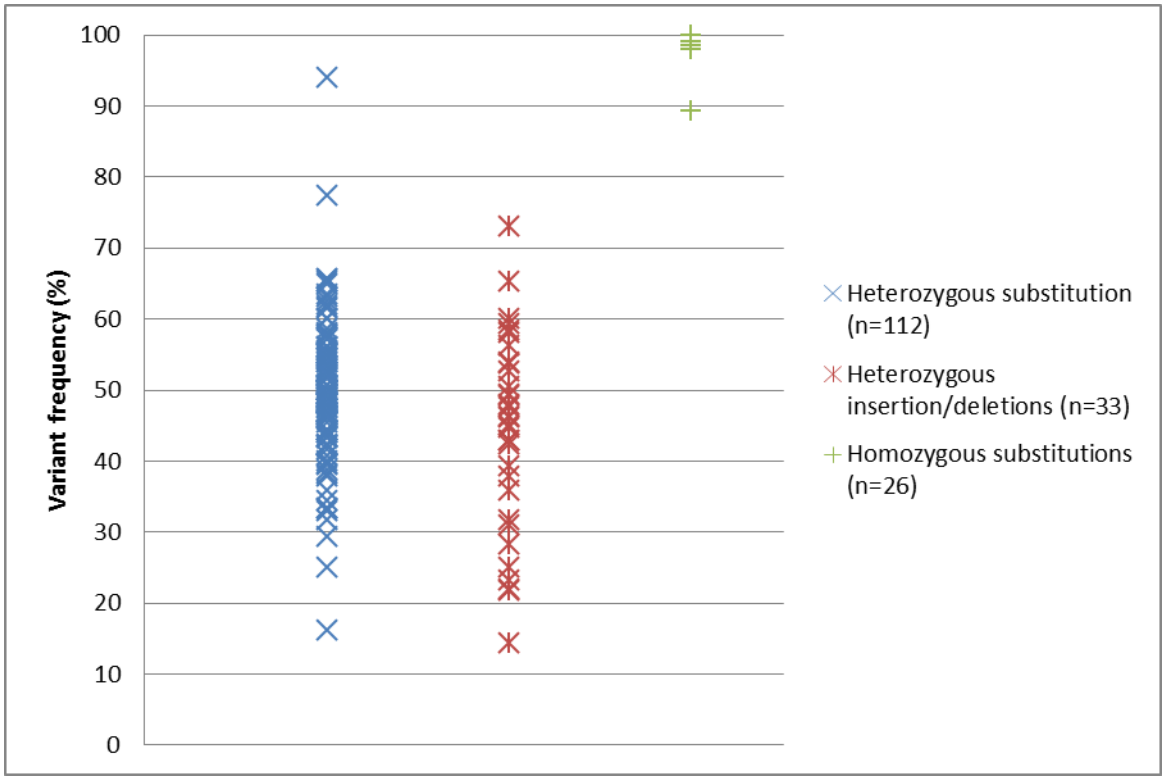


Figure 3.3 Variant frequency observed for the variants input into the PHEO/PGL/HNPGL panel during validation that met the >x30 coverage threshold (n=171)

3.4.2.4 RCC panel validation

3.4.2.4.1 RCC panel sensitivity

For the RCC panel validation, 42 DNA samples were run, comprising 106 variants. 3 of which did not meet the coverage threshold of ≥30 reads. Of the variants examined, 50 were unique (see Table 3.5) and 49 had reads which met the required coverage threshold. All variants were detected, providing a sensitivity of 100% (95% CI: 92.68% to 100.00%).

Table 3.5 Unique variants used to validate the RCC panel

Gene	Exon/ Intron	Nucleotide change	Protein change	Variant type
FH	2	c.268-28T>A		Substitution
FH	2	c.268-24A>T		Substitution
FH	2	c.268-22A>T		Substitution
FH	5	c.698G>A	p.Arg233His	Substitution
FH	6	c.904+47G>A		Substitution
FH	7	c.1108+98C>A		Substitution
FH	7	c.1108+109G>A		Substitution
FH	7	c.927G>A	p.=	Substitution
FH	9	c.1347delG	p.Met449Ilefs*5	Insertion/ Deletion
FH	9	c.1391-1G>T		Substitution
FLCN	2	c.347dupA	p.Leu117Alafs*16	Insertion/ Deletion
FLCN	3	c.469_471delTTC	p.Phe157del	Insertion/ Deletion
FLCN	3	c.583G>T	p.Gly195*	Substitution
FLCN	4	c.240delC	p.Asp80Glufs*50	Insertion/ Deletion
FLCN	5	c.340dupC	p.His114Profs*19	Insertion/ Deletion
FLCN	5	c.396+59T>C		Substitution
FLCN	5	c.397-14C>T		Substitution
FLCN	6	c.494delG	p.Gly165Alafs*12	Insertion/ Deletion
FLCN	6	c.890_893delAAAG	p.Glu297Alafs*25	Insertion/ Deletion
FLCN	8	c.871+226G>A		Substitution
FLCN	8	c.871+36G>A		Substitution
FLCN	9	c.1062+6C>T		Substitution
FLCN	9	c.1062+47G>A		Substitution
FLCN	10	c.1076delC	p.Pro359Leufs*16	Insertion/ Deletion
FLCN	10	c.1176+1G>C		Substitution
FLCN	10	c.1176+39G>A		Substitution
FLCN	10	c.1177-5_1177-3delCTC		Insertion/ Deletion
FLCN	11	c.1285dupC	p.His429Profs*27	Insertion/ Deletion
FLCN	12	c.1301-59C>T		Substitution
FLCN	12	c.1301-38G>A		Substitution

Gene	Exon/ Intron	Nucleotide change	Protein change	Variant type
FLCN	12	c.1318_1334dup17	p.Leu449Glnfs*25	Insertion/ Deletion
FLCN	12	c.1367_1398del32	p.Asp456Glyfs*19	Insertion/ Deletion
MET	17	c.3577-75T>C		Substitution
SDHB	1	c.72+1G>T		Substitution
SDHB	2	c.200+104A>G		Substitution
SDHB	2	c.200+33G>A		Substitution
SDHB	2	c.79C>T	p.Arg27*	Substitution
SDHB	3	c.268C>T	p.Arg90*	Substitution
SDHB	3	c.286+169A>G		Substitution
SDHB	4	c.380T>G	p.Ile127Ser	Substitution
SDHB	4	c.423+1G>A		Substitution
SDHB	6	c.590C>G	p.Pro197Arg	Substitution
SDHB	6	c.600G>T	p.Trp200Cys	Substitution
VHL	1	c.240T>G	p.Ser80Arg	Substitution
VHL	2	c.350G>T	p.Trp117Leu	Substitution
VHL	2	c.457delC	pLeu153Cysfs*6	Insertion/ Deletion
VHL	2	c.463+108T>G		Substitution
VHL	3	c.491A>G	p.Glu164Arg	Substitution
VHL	3	c.525C>A	p.Tyr175*	Substitution
VHL	3	c.568_567dupGA	p.Asp190Glufs*13	Insertion/ Deletion

3.4.2.4.2 RCC panel specificity

The same filtering system designed for the PHEO/PGL/HNPGL panel was used to limit false variant calls resulting from errors in determining the number of bases within homopolymer tracts. Within the specified region of interest, the genes analysed on this panel do not have many homopolymeric stretches. In the validation phase, this meant over 80% fewer deletion/insertion calls required additional analysis and of those flagged as requiring additional investigation 81% were true variants. No true variants were missed.

3.4.2.4.3 RCC panel variant frequency

Figure 3.4 provides details of the variant frequency observed for the 103 variants detected by NextGENe. Again, the average variant frequency was close to that expected for all variant types. The heterozygous substitution variants were seen in an average of 46.8% of reads (range: 24.4-67.9%) and the heterozygous insertion/deletion variants in an average of 45.8% (range: 22.9-66.3%). The homozygous substitutions were seen at an average of 99.7% (range: 95.6-100.0%). Again for this panel, homozygous substitutions were not always identified in 100% of reads (n=5/27).

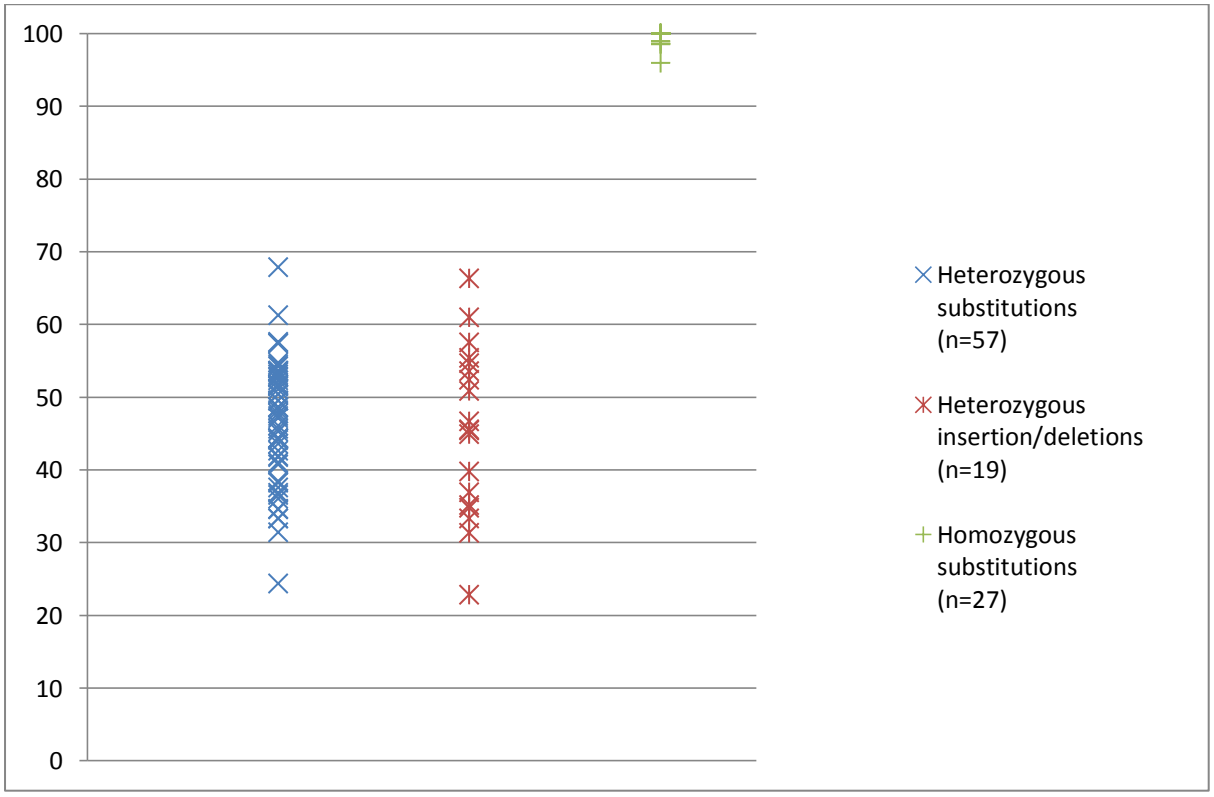


Figure 3.4 Variant frequency observed for the variants input into the RCC panel during validation that met the >30x coverage threshold (n=130)

3.4.3 Workflow establishment

After optimisation and validation of each panel a full workflow was determined. This was common for both panels.

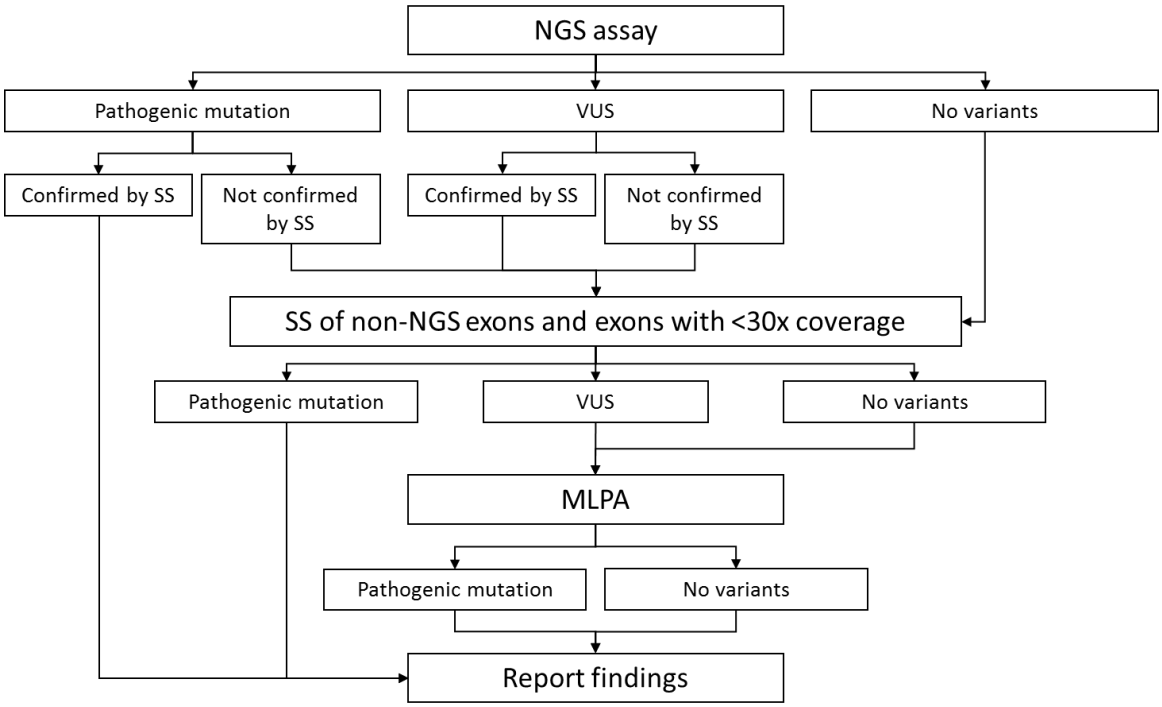


Figure 3.5 Panel workflow, common to both the PHEO/PGL/HNPGL and RCC panels

For the PHEO/PGL/HNPGL panel the non-NGS exons were *MAX* exon 1, *SDHA* exon 1, *SDHAF2* exon 1 and *TMEM127* exon 2. The MLPA probemixes used were P226 SDH, which includes *SDHA*, *SDHC* and *SDHD* and P016 VHL which includes the *VHL* gene and flanking region.

For the RCC panel, *FH* exon 1 and *FH* exon 9 were non-NGS exons. The MLPA probemixes used were P256 FLCN, which includes the *FLCN* gene, P226 SDH, which includes *SDHB*, *SDHC* and *SDHD* and P016 VHL which includes the *VHL* gene and flanking region.

3.4.4 Prospective patient analysis

3.4.4.1 Prospective patient analysis PHEO/PGL

The second phase involved prospective analysis of 120 patient samples using the PHEO/PGL/HNPGL workflow. The mean number of passed filter reads was ~108, 000 per run, higher than the average of the validation runs (~82, 000) and increasing amplicon coverage. This was likely to be the result of a Roche software upgrade during that period, in combination with increased user experience. The complete genetic workflow (Figure 3.5) identified 44 variants (excluding benign polymorphisms of no clinical significance), of which 36 were unique. 12 were classed as pathogenic, 3 as likely pathogenic and 21 as a variant of unknown significance (VUS) (5 patients had multiple variants). The variants have been summarized in Table 3.6. An example of a pathogenic nonsense mutation detected by NGS and confirmed by Sanger sequencing is shown in Figure 3.6.

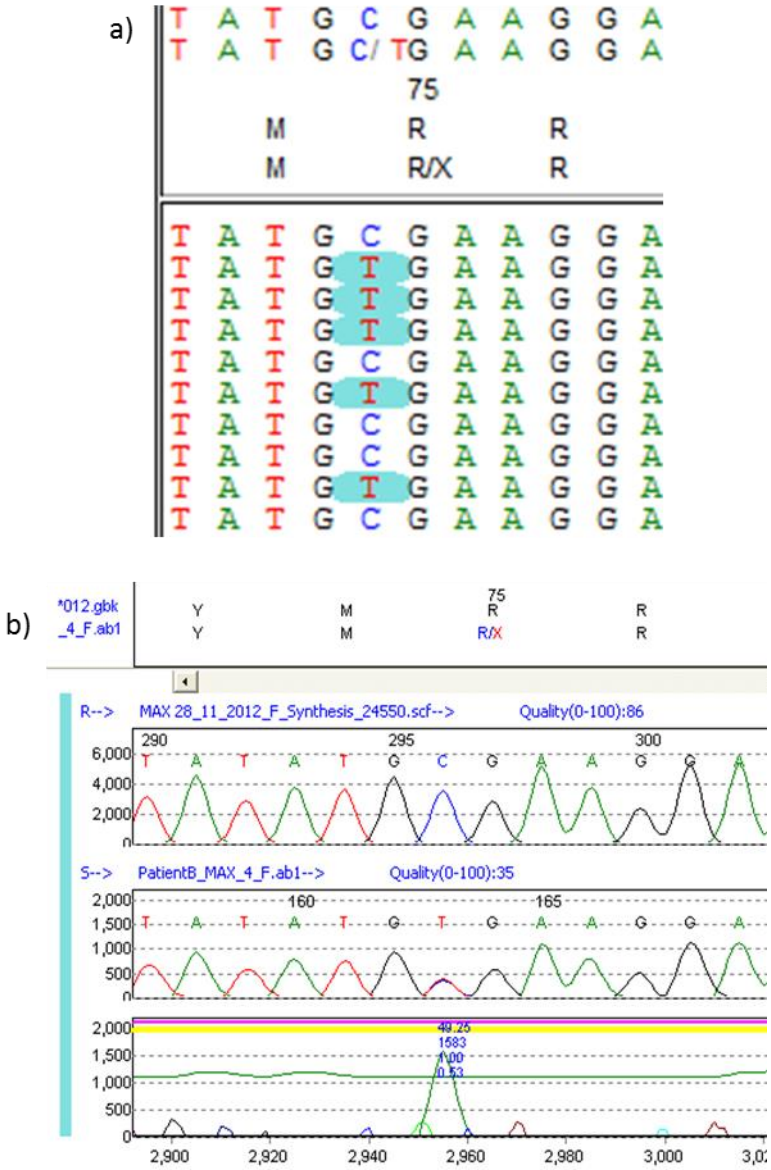


Figure 3.6 An example of a pathogenic mutation in the *MAX* gene, c.[223C>T]; [=] p.(Arg75*), identified by NGS and confirmed by Sanger sequencing
a) NGS read pile up of ten example reads visualised by NextGENe; b) Sanger sequencing forward electropherogram visualised by Mutation Surveyor.

As might be expected, the detection rate for novel pathogenic variants was inversely related to the extent of previous molecular genetic analysis. One or more genes had been sequenced previously in 35% of patients, of whom six patients had a VUS identified (all

confirmed by the NGS panel). Ten additional variants in seven patients (three of whom had a previously diagnosed VUS) were identified. All 10 additional variants were identified in genes that had not previously been analyzed and three were considered likely to be pathogenic.

No previous genetic analysis had been performed in 65% of patients and a total of 28 variants were identified in 27 individuals. Of these, 46% were considered to be pathogenic mutations (10 in *SDHB, SDHD* and *VHL* and 3 in genes that were not previously part of the routine diagnostic service). In addition, 3 likely pathogenic changes and 12 VUSs were identified (two of which were identified in one patient). In summary, a variant of interest was identified in 34.6% of new referrals and a definite pathogenic mutation in 16.6% (see Table 3.6).

3.4.4.2 Prospective patient analysis RCC

For the RCC panel 119 patients were analysed prospectively. The complete RCC genetic workflow identified 12 unique variants (2 were seen twice), 5 were pathogenic and 7 were VUSs, with 2 VUSs seen in 1 patient (see Table 3.7). This equates to a variant detection rate of 11/119 (9.2%) and a genetic diagnosis confirmed in 5/119 patients (4.2%).

As with the PHEO/PGL/HNPGL panel the detection rate was affected by the amount of previous analysis performed, one or more genes has been analysed previously in 23/119 (19.3%) of patients, of whom 1 had previously had a VUS identified (detected by NGS). No additional variants were identified in any of the previously analysed patients.

A variant of interest was identified in 14/119 (11.8%) patients and a pathogenic mutation identified in 5/119 (4.2%).

Table 3.6 Variants identified during prospective PHEO/PGL/HNPGL panel analysis.
HNPGL = head and neck paraganglioma, PGL = paraganglioma (abdominal), PHEO = phaeochromocytoma (adrenal), VUS = variant of uncertain significance.

Gene	Exon/ Intron	Variant type	Variant (coding DNA)	Variant (protein)	Times observed	Variant class	Tumour type
MAX	4	Nonsense	c.[223C>T];[=]	p.(Arg75*)	1	Pathogenic	PHEO
MAX	5	Missense	c.[425C>T];[=]	p.(Ser142Leu)	1	VUS	PHEO
SDHA	2	Nonsense	c.[91C>T];[=]	p.(Arg31*)	2	Likely pathogenic	1. HNPGL 2. PGL
SDHA	2	Missense	c.[136A>G];[=]	p.(Lys46Glu)	1	VUS	PGL
SDHA	5	Synonymous	c.[549C>T];[=]	p.(=)	1	VUS	HNPGL
SDHA	7	Synonymous	c.[822C>T];[=]	p.(=)	1	VUS	HNPGL
SDHA	8	Missense	c.[923C>T];[=]	p.(Thr308Met)	1	VUS	PGL
SDHA	8	Synonymous	c.[1002G>A];[=]	p.(=)	1	VUS	HNPGL
SDHA	10	Missense	c.[1273G>A];[=]	p.(Val425Met)	1	VUS	PHEO/PGL
SDHA	10	Frameshift	c.[1338delA];[=]	p.(His447Metfs*23)	1	VUS	HNPGL
SDHA	12	Synonymous	c.[1623G>A];[=]	p.(=)	1	VUS	HNPGL
SDHA	13	Missense	c.[1753C>T];[=]	p.(Arg585Trp)	1	Likely pathogenic	HNPGL
SDHA	13	Synonymous	c.[1776T>C];[=]	p.(=)	1	VUS	PGL
SDHAF2	3	Missense	c.[319C>T];[=]	p.(Arg107Cys)	1	VUS	PGL
SDHB	1	Splicing	c.[72+1G>T];[=]		1	Pathogenic	PHEO
SDHB	1	Deletion	c.[-?_200+?del];[=]	p.0?	1	Pathogenic	PGL
SDHB	2	Deletion	c.[73-?_200+?del];[=]		1	Pathogenic	PGL
SDHB	2	Missense	c.[118A>G];[=]	p.(Lys40Glu)	1	Pathogenic	PGL
SDHB	4	Splicing	c.[423+1G>A];[=]		1	Pathogenic	PHEO

Gene	Exon/ Intron	Variant type	Variant (coding DNA)	Variant (protein)	Times observed	Variant class	Tumour type
SDHB	5	Missense	c.[487T>C];[=]	p.(Ser163Pro)	5	VUS	1. PHEO 2. PHEO 3. PHEO 4. PHEO 5. HNPGL
SDHB	6	Missense	c.[587G>A];[=]	p.(Cys196Tyr)	1	Pathogenic	HNPGL
SDHB	7	Missense	c.[725G>A];[=]	p.(Arg242His)	1	Pathogenic	HNPGL
SDHB	8	Frameshift	c.[770dupT];[=]	p.(Asn258Glufs*17)	1	Likely pathogenic	PHEO
SDHB	2-7	Deletion	c.[73-?_765+?del];[=]		1	Pathogenic	PGL
SDHC	1	5'UTR	c.[-118_-117delAG];[=]		1	VUS	HNPGL
SDHC	2	Splicing	c.[77+2dupT];[=]		1	Pathogenic	HNPGL
SDHC	3	Missense	c.[148C>T];[=]	p.(Arg50Cys)	1	VUS	HNPGL
SDHC	4	Missense	c.[214C>T];[=]	p.(Arg72Cys)	1	VUS	HNPGL
SDHC	5	Missense	c.[380A>G];[=]	p.(His127Arg)	1	VUS	HNPGL
SDHD	1	Missense	c.[34G>A];[=]	p.(Gly12Ser)	3	VUS	1. HNPGL 2. PHEO 3. HNPGL
SDHD	3	Missense	c.[242C>T];[=]	p.(Pro81Leu)	2	Pathogenic	1. HNPGL 2. HNPGL
SDHD	3	Synonymous	c.[312C>T];[=]	p.(=)	1	VUS	PGL
TMEM127	3	Missense	c.[268G>A];[=]	p.(Val90Met)	1	VUS	PGL
TMEM127	3	Frameshift	c.[512delTinsGCC];[=]	p.(Val171Glyfs*137)	1	Pathogenic	HNPGL
TMEM127	4	Synonymous	c.[411T>A];[=]	p.(=)	1	VUS	HNPGL
TMEM127	4	Synonymous	c.[534C>T];[=]	p.(=)	1	VUS	PHEO

Table 3.7 Variants identified during prospective RCC panel analysis.

VUS = variant of uncertain significance.

Gene	Exon/Intron	Variant type	Variant (coding DNA)	Variant (protein)	Times observed	Variant class
FLCN	7	Missense	c.[715C>T];[=]	p.(Arg239Cys)	2	VUS
FLCN	11	Frameshift	c.[1285dupC];[=]	p.(His429Profs*27)	1	Pathogenic
FLCN	12	Missense	c.[1333G>A];[=]	p.(Ala445Thr)	1	VUS
FLCN	12	Frameshift	c.[1333_1337dupGCACG];[=]	p.(Ser447Hisfs*23)	1	Pathogenic
FLCN	3'UTR	3'UTR	c.[*4G>A];[=]		1	VUS
MET	17	Missense	c.[3356G>C];[=]	p.(Gly1119Ala)	1	VUS
SDHB	1	Missense	c.[32G>A];[=]	p.(Arg11His)	1	VUS
SDHB	5	Missense	c.[487T>C];[=]	p.(Ser163Pro)	2	VUS
SDHB	7	Missense	c.[724C>T];[=]	p.(Arg242Cys)	1	Pathogenic
VHL	1	Missense	c.[233A>G];[=]	p.(Asn78Ser)	1	Pathogenic
VHL	3	Missense	c.[629G>A];[=]	p.(Arg210Gln)	1	VUS
VHL	3	Splice-site	c.[464-2A>G];[=]		1	Pathogenic

3.5 Discussion

The hypothesis for this chapter was that NGS could facilitate a transition from targeted, sequential, analysis of individual PHEO/PGL/HNPGL or RCC genes in selected high risk individuals to a strategy of simultaneously testing multiple predisposition genes in all at-risk individuals. To determine if this strategy was likely to be a feasible innovation for clinical practice it was necessary to establish a) whether NGS could be used as a reliable alternative to Sanger sequencing to interrogate all relevant parts of the genes of interest and b) the sensitivity and specificity by validation of the NGS test.

The AA was found to be an effective system to produce an amplicon library for the GS Junior and has the advantage that additional amplicons could be included as further genes applicable to either the PHEO/PGL/HNPGL or RCC phenotype are identified. However the validation processes identified some exons unsuitable for analysis in this workflow. In particular those with a GC-content greater than 66% tended to produce poor results and this should be considered when designing assays using this system. Also, as in the case of *FH* exon 9, very occasionally amplicons will not work in the assay for no determinable reason.

As these were novel assays, validation for diagnostic use was essential. For the PHEO/PGL/HNPGL panel, analytical sensitivity was experimentally determined to be 98.7%. Only 1/77 unique variants was not detected during phase one. However, this undetected variant was identified on five other occasions (although in a lower average percentage of reads than might be expected for a heterozygous change). This was likely to be the result of sequence context, as the technology appeared to have difficulty sequencing this region. The variant in question was a benign polymorphism and the failure to detect it was not of

clinical significance but it is important to consider that in other circumstances a pathogenic mutation on an underrepresented allele may be overlooked. Also, it is important to note that non-homopolymer repeat tracts, such as a TTC repeat, can also lead to sequencing errors when using Roche NGS technology. For the RCC genes the analytical sensitivity was experimentally determined to be 100%.

There is no consensus regarding the minimum coverage and variant filtering threshold (VFT) required to reliably detect a heterozygous variant. Though de Leeneer *et. al.* theoretically calculated that a 15% VFT with 30-fold coverage would allow detection of 99.995% of heterozygous variants, this assumes that neither allele is preferentially amplified or sequenced (98). If this is not the case, then higher coverage would be required. Several groups have used a similar enrichment strategy and sequencing chemistry, and set locally derived limits to detect heterozygous mutations. For example Hollants *et al.* (99) proposed that a VFT of 20% and a minimum coverage of 25-fold was sufficient to detect heterozygotes in familial hypercholesterolemia genes, whereas for BRCA1 and BRCA2 screening Michils *et. al.* suggested that 10% VFT and 27-fold coverage was required to detect all heterozygous changes (100). Our findings suggest a 15% VFT with 30-fold coverage is required to optimize variant detection whilst attenuating the concomitant increase in false positive results, for this combination of methodologies.

It is well established that homopolymer tract sequencing errors are problematic when using the Roche 454 chemistry (101). To reduce the number of variants requiring Sanger sequencing confirmation, the regions analysed were reduced to comprise the coding sequence $\pm$5 bp of flanking intronic sequence for all genes. Additionally, a variant

comparison strategy was implemented to simultaneously interrogate all variants in all patient DNAs on a single NGS run. Any deletion/duplication called was compared to the data from the same base for all samples sequenced on that run. Only variants for which the percentage deletion or duplication call was greater than 2 standard deviations from the mean were confirmed by Sanger sequencing. When these filters were applied to the validation data, all known mutations were detected and the number of false positive calls was dramatically reduced. Not all NGS chemistries have inherent issues with homopolymer tract sequencing and other platforms such as the Illumina MiSeq would offer some advantages in this respect. However this was not an available option for this project at the time.

For the PHEO/PGL/HNPGL panel, the largest deletion identified during the validation phase was 7 bp in *SDHD* exon 1 and the largest duplication 13 bp in *SDHB* exon 7. For the RCC panel the largest deletion and duplication observed in the validation phase were both in FLCN exon 12 and were 32 bp and 17 bp in size, respectively. The ability to identify whole exon deletions and duplications was not thoroughly examined given that the enrichments were PCR based and therefore highly unlikely to produce reliable data. Additionally a dosage analysis tool was not available in the commercial software used for analysis. Brief attempts to use normalised average exon read numbers to detect whole exon copy number changes reinforced that theory (data not included).

As reported previously for similar assays (100;102;103), the variant frequency was variable for both heterozygous and homozygous changes. For homozygous changes, the lack of 100% variant calls can be explained by low level sequencing errors. However for heterozygous

changes, the large variability may reflect random sampling of the PCR product. This is exemplified by a heterozygous substitution for which the same DNA was run on three occasions and gave variant frequencies of 32.91, 53.03%, 94.03%. Investigation of the DNA suggested that the variation may have been exacerbated by mild DNA degradation. For a direct amplicon-based sequencing strategy, duplicate reads cannot be identified and removed (as would occur when performing a method using fragmentation), which may lead to bias for one allele, and hence the requirement for a lower VFT. An alternative target enrichment methodology could be trialed to determine whether a capture and fragmentation approach would lead to less allelic bias.

After establishing a diagnostic testing workflow for both PHEO/PGL/HNPGL and RCC, a prospective analysis of samples referred for diagnostic testing was undertaken, with use of the panels upon clinician request. It is interesting to note the difference in the pathogenic mutation identification rate for the PHEO/PGL/HNPGL panel vs the RCC panel (16.6% vs 4.2%). The low RCC rate is likely to be the result of referral bias. It appeared that those patients with a likely diagnosis of a specific condition were often not referred for panel testing but rather for the specific gene of interest, meaning that those samples referred for the panel test were more likely to be from patients where the clinician was less certain of the likelihood that it was an inherited RCC or where the case was complex. Two interesting cases are described briefly in section 3.6. In contrast PHEO/PGL/HNPGL are more rarely identified as part of a clinically clear syndrome and there are more candidate genes for each tumour subtype (PHEO, PGL and HNPGL). Therefore the PHEO/PGL/HNPGL panel was used for the vast majority of referred cases.

There were two major advantages for the NGS workflows. The first was that each group of genes was analysed simultaneously whereas conventional sequential molecular genetic testing strategies analyse one gene and then proceed to testing another (if no mutation is detected). Thus, the workflows decreased the time taken to find a pathogenic mutation or determine that no pathogenic mutation could be identified in the analysed genes. In practice, this would enable predictive testing to be made available to at risk family members in a timelier manner. National guidelines suggest that testing for an unknown mutation in a large gene should be performed within 40 working days and that an NGS panel test should be performed within 80 working days. The second major advantage was the reduced cost per gene analysed, which was most diminished for the PHEO/PGL/HNPGL panel. The diagnostic screening workflow was locally priced by the NHS diagnostic laboratory at £500 per sample (for NHS diagnostic referrals) to test 9 PPGL/HNPGL genes (*MAX, RET, SDHA, SDHB, SDHC, SDHD, SDHAF2, TMEM127 and VHL*), or ~£56 per gene. Compared to the local charge for testing *VHL, RET, SDHB* and *SDHD* by conventional sequential (Sanger) sequencing technology of ~£1800 (104). This equates to a 70% cost reduction for users with a 125% increase in the number of genes examined. The RCC panel was locally priced at £550 per sample for analysis of 5 genes or ~£110 per gene, compared to testing of any of the RCC genes alone which were charged at between £400 and £800 depending on the number of exons in the gene. This meant that there was minimal cost difference for the analysis of an additional 4 genes.

An inevitable consequence of testing more genes is the identification of additional VUSs. For the PHEO/PGL/HNPGL panel, it was found that 61% of variants detected were VUSs and for the RCC panel, 64% of variants detected were classed as VUSs. In a diagnostic laboratory,

accurate variant classification is essential as it impacts directly on patient care and family management. In the short term, the identification of a VUS can cause uncertainty for the patient and additional workload for the laboratory staff and clinicians. For the PHEO/PGL/HNPGL panel, 51.9% of the VUSs (but only 15.4% of pathogenic mutations) were in *MAX, SDHA, SDHAF2* or *TMEM127*, genes not analyzed previously by the diagnostic laboratory. These genes were all relatively newly identified and thus there was a paucity of references to them in the literature. It is anticipated that as research and diagnostic testing of these genes expands, more VUSs will be assigned to pathogenic or benign classes. Classification will be further aided by the large population studies which are now being performed, such as the NHLBI GO Exome Sequencing Project. Nevertheless, it is anticipated that some VUSs will remain difficult to classify and will require in depth clinical, *in silico* and *in vitro* investigation. Testing multiple genes in parallel can aid VUS classification such that if VUS is identified with a known pathogenic mutation in the same or another gene then, unless there is digenic inheritance, the likelihood of the VUS being pathogenic is reduced (though genetic information should always be interpreted in the context of clinical and pathological findings, including immunohistochemical studies).

In theory this assay could be applied to the analysis of fresh frozen tumour samples but this would require the analysis of fewer samples so that coverage could be increased to detect clonal variants. Additionally, another set of primers could be designed with shorter amplicon lengths in order to analyse FFPE material.

3.6 Case Studies

3.6.1 RCC case 1

The patient had an oncocytoma diagnosed at 40 years. They also had two skin tags and two lesions on their trunk. The clinician queried a diagnosis of BHD. The panel identified a pathogenic missense mutation in *SDHB*, c.724C>T p.(Arg242Cys). A different pathogenic mutation in SDHB has previously been reported in association with oncocytoma (105).

3.6.2 RCC case 2

The female patient is one of eight siblings and has no family history of disease. She had bilateral renal cancer diagnosed at 36 years of age. One renal tumour was papillary type and the other was clear cell type. The clinician queried HLRCC. The panel identified a *FLCN* frameshift mutation, c.1333_1337dupGCACG p.(Ser447Hisfs*23). Retrospective analysis of the patient identified no other symptoms consistent with a BHD phenotype. For example there were no skin abnormalities consistent with fibrofolliculomas. This finding allowed this patient to be counselled regarding her risk of the other clinical features of BHD and her large family to have predictive genetic analysis.

3.7 Conclusion

In summary, for this study two novel diagnostic assays were developed and validated. Both panels allowed increased detection of germline mutations in patients with PHEO/PGL/HNPGL and RCC at a lower cost per gene and reduced processing time compared to conventional sequential (Sanger-based) molecular genetic analysis. For the PHEO/PGL/HNPGL panel, in a cohort of 120 prospectively analyzed patients, 10 pathogenic germline mutations were identified in the genes analysed in the previous "standard" mutation analysis strategy (*SDHB, SDHD, RET* and *VHL*) and a further 3 mutations by analysing the less frequently interrogated PHEO/PGL/HNPGL genes, a 30% increase in diagnostic yield compared to the previous strategy. For the RCC panel, the case studies show examples of expedited molecular genetic diagnosis of pathogenic mutations in genes that may not have been otherwise requested by the referring clinician. Longer term it is predicted that better classification of potential VUSs will further enhance the utility of assays similar to this and enable the advantages of precise molecular diagnosis to be offered to larger cohorts of patients.

Chapter 4 Development of an enhanced next generation sequencing (NGS) strategy for improved genetic diagnosis of von Hippel-Lindau (VHL) disease

4.1 Introduction

4.1.1 von Hippel-Lindau (VHL) disease and its name

von Hippel-Lindau (VHL) disease is an autosomal dominant inherited condition which causes multiple benign and malignant tumours including clear cell renal carcinoma (ccRCCs), retinal and central nervous system (CNS) haemangioblastomas (HABs), phaeochromocytomas (PHEOs), pancreatic neuroendocrine tumours (106) and some additional rarer tumours. It is thought to affect approximately 1 in 36, 000 individuals (107). VHL was named after Eugen von Hippel, who described a family with retinal angiomas in 1904 (108), and Arvid Lindau, who described the cerebellar and spinal HABs in 1927 (109), but the term did not become commonly used until the 1970s (106). Prior to the advent of molecular diagnostics a clinical diagnosis of VHL was given when a patient had two relevant tumours without a family history of disease or one tumour with a family history (110).

4.1.2 *VHL* gene identification and tumour suppressor activity

In 1993, positional cloning identified the *VHL* gene at 3p25-26 (111). It is a tumour suppressor gene. VHL disease-related tumour development results from the combination of an inherited mutation on one allele of *VHL* and inactivation or loss of the second allele, conforming to Knudson's two hit hypothesis (26). This was initially demonstrated when loss of heterozygosity (LOH) analysis found that inactivation of both *VHL* alleles was required for tumour formation to occur, both for VHL-related tumours and sporadic ccRCC (112-114). Biallelic loss of VHL protein (pVHL) is found in the majority of sporadic ccRCC (115;116). In addition, it was shown that when pVHL was reintroduced into *VHL*-null ccRCC cells it

supressed their ability to form tumours in nude mice (117) and restored their capacity to exit the cell cycle and become quiescent upon serum withdrawal, which suggests a role for pVHL in regulation of cell cycle exit in the kidney (118).

4.1.3 VHL phenotypic heterogeneity

4.1.3.1 VHL disease

VHL disease exhibits phenotypic heterogeneity (119;120) and is classified into various types dependent on the clinical features observed in the patient. Type 1 and Type 2 are differentiated by the absence or presence of PHEO respectively (119)**.** Type 2 is then further subdivided, depending on the presence or absence of RCC and other tumours, as described in Table 4.1 (121). Whilst the division can aid genotype-phenotype analysis, families have been shown to move between the types (106) and this classification is not used widely for clinical management.

Table 4.1 Genotype-phenotype correlations in VHL disease
(Adapted from (106)).

VHL disease subtype	Clinical manifestations	Type of *VHL* gene mutations observed
1	ccRCC, HAB	Deletions (not including upstream *BRK1*), nonsense, frameshift, missense
1B	HAB	Deletion including *BRK1*
2A	HAB, PHEO	Missense
2B	ccRCC, HAB, PHEO	Missense
2C	PHEO	Missense

4.1.3.2 Chuvash polycythaemia (CSP)

Mutations in the *VHL* gene are also known to cause Chuvash polycythaemia (CSP), the first inherited disorder of oxygen sensing to be reported (122). The CSP phenotype consists of varicose veins, vertebral haemangiomas, elevated serum VEGF concentration and increased risk of stroke (123). Erythropoietin is high compared to haemoglobin in the majority of patients (123). In contrast to VHL disease, which is autosomal dominant and caused by a large variety of heterozygous germline mutations, CSP is autosomal recessive and caused by specific *VHL* germline missense mutations when present homozygously or compound heterozygously (122;124-127). The c.598C>T p.(Arg200Trp) mutation (also known as R200W) is the most common CSP mutation and is endemic to the Chuvash Autonomous Republic of the Russian Federation, hence the disease name (128). In the literature there are cases of patients with CSP and an apparently single heterozygous *VHL* mutation (128). The vast majority of patients with CSP do not develop tumours, but there have been two reports of patients with isolated HAB (106) and one compound heterozygous patient who presented with polycythaemia aged 7 and PHEO in his 30s (129). A knock-in R200W transgenic mouse shows polycythaemia without tumour formation (130) and it is believed that heterozygous carriers of c.598C>T p.(Arg200Trp) do not have an increased risk of cancer (131;132). Additionally, polycythaemia is rarely observed in individuals with VHL disease and then only as a component of paraneoplastic syndrome in those with a tumour (133;134).

4.1.4 *VHL* mRNA isoforms

The *VHL* gene is composed of three exons. There are two mRNA isoforms: 1) NM_000551, which is the longer and includes all 3 exons and 2) NM_198156, which lacks exon 2, an in-frame coding exon (135-137). Given that deletion of exon 2 is a pathogenic mutation which leads to VHL Type 1, the isoform lacking exon 2 cannot be sufficient for pVHL function (138;139).

4.1.5 The VHL protein (pVHL) and its function

There are two isoforms of pVHL resulting from the longer mRNA transcript. The larger consists of 213 amino acids and encodes a 30kDa protein named $pVHL_{30}$ (140). The smaller isoform consists of 160 amino acids and encodes a 19kDa protein named $pVHL_{19}$ (141;142). $pVHL_{19}$ is the result of an internal methionine at amino acid position 54; when it is used the protein does not include an acidic pentamer repeat domain. It has been shown that $pVHL_{30}$ co-localises principally with cytoplasmic microtubules, whereas $pVHL_{19}$ localises largely in the nucleus (143). The smaller isoform is thought to be predominant in multiple tissues (106). Early functional studies suggested the two isoforms have equivalent effects in assays (141) and both isoforms possess tumour suppressor activity *in vivo* (142). There is strong evolutionary conservation over the majority of $pVHL_{19}$, however the first 53 amino acids of $pVHL_{30}$ do not exhibit such strong conservation and there is an extreme paucity of pathogenic familial mutations reported in that region compared to the rest of the gene (144).

pVHL comprises two structural domains, the alpha and beta domains. The alpha domain is mainly α-helices and acts as a binding site. The beta domain comprises β-sheets, is

approximately 100 residues in size and is involved in substrate recognition (145). pVHL forms a ternary complex with the elongin C and elongin B proteins (146-149), named the VCB complex, which is critical for pVHL's function. Figure 4.1 depicts the structure (149). The VCB complex nucleates another complex containing CUL2 and RBX1 to create the VCB-CR complex (150-152). The formation of the VCB complex stabilises its component proteins (153) and makes it resistant to proteasomal degradation. However if pVHL contains a mutation which disrupts elongin binding, it becomes unstable and rapidly degraded by the proteasome (146-148;150).

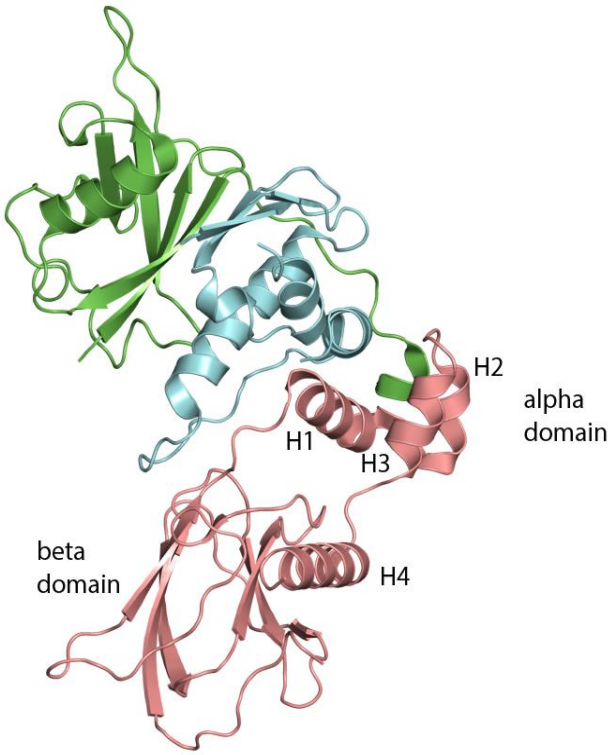


Figure 4.1 The VCB complex
The VCB complex has two interfaces, the first between VHL (pink) and elongin C (blue) and the second between elongin C and elongin B (green). (Adapted from (106)).

4.1.5.1 Hypoxia-inducible factor (HIF)-dependant VHL functions

pVHL is known to have a variety of functions but the best understood is its role in oxygen sensing, as shown in Figure 4.2. In VHL disease, this pathway is thought to contribute to tumourigenesis because under normoxic conditions pVHL is necessary to switch off HIF action. Without any functional pVHL, inappropriate activation of HIF downstream target genes occurs despite normal oxygen levels. Many of the pathogenic *VHL* mutations are predicted, or have been demonstrated, to impair interaction between pVHL and HIF (154-158).

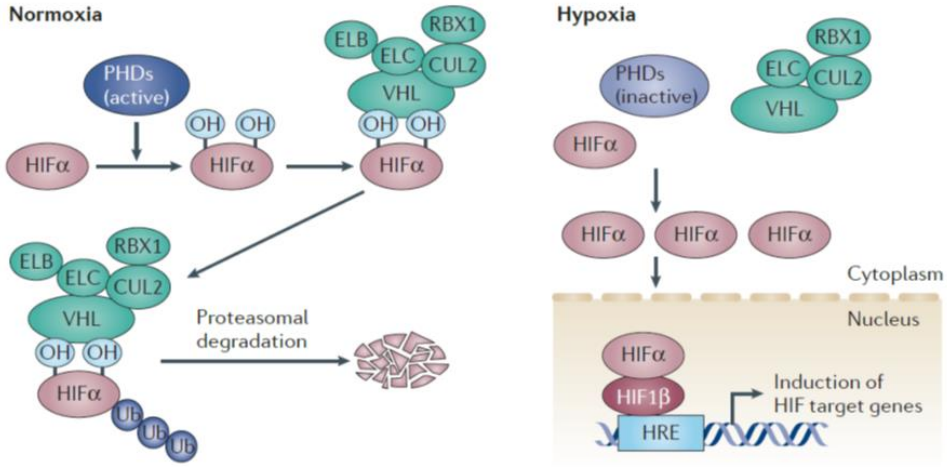


Figure 4.2 Schematic depicting pVHL function in normoxia and hypoxia
In the presence of oxygen (normoxia), HIFα proteins bind exclusively to the beta domain of pVHL (154;155). The binding of HIFα by pVHL is dependent upon its hydroxylation by prolyl hydroxylases 1-3 (PHD1-3) at two conserved proline residues (Pro564 and Pro402 for HIF1α), also known as the oxygen-dependant degradation domain (123). PHD1-3 only function under normoxic conditions as they require oxygen as a co-substrate (155;159-162). The prolyl-hydroxylation of HIFα allows it to be recognised and ubiquitinated by the VCB-CR complex and thus recognised and degraded by the cellular proteasome. In contrast, under hypoxic physiological conditions, or when functional pVHL is lost, HIFα is not degraded and therefore accumulates and forms heterodimers with HIFβ, which translocates to the nucleus binds to hypoxia response elements(163). It is thought that there are more than 800 HIF target genes (164;165); many of which promote acute or chronic hypoxia adaption(166). In addition, HIFs are thought to bind and transactivate genes, encoding microRNAs (167) and chromatin modifying enzymes (164;168-170), and thus indirectly regulate genes (Adapted from (106)).

Three isoforms of HIFα have been identified, HIF1α, HIF2α and HIF3α. Knowledge of HIF3α is limited, but more extensive work has been done on HIF1α and HIF2α. HIF1α is ubiquitously expressed whereas HIF2α is mainly expressed in endothelial, lung, renal and hepatic cells (171). In hypoxic conditions, both HIF1α and HIF2α activate transcription of target genes by binding to the same hypoxia-response element, following dimerization with HIF1β (172). They have been shown to promote transcription of different genes with HIF1α promoting genes involved in the glycolytic pathway and HIF2α promoting growth and angiogenesis genes (173-175) . It is thought that HIF2α is the key driver of renal cancer progression (reviewed in (172)).

4.1.5.2 HIF independent functions

pVHL also has HIF-independent functions and it is these functions that are likely to explain the VHL type 2C phenotype, in which only PHEOs are observed and the mutations have been shown to maintain their ability to downregulate HIF (121). The molecular mechanism central to the pathogenesis of *VHL*-associated CSP is unclear, but two theories have been presented. The first suggests that these variants only mildly affect oxygen sensing by only affecting a subset of HIF activated genes and that a greater level of dysfunction is required for tumour formation, whereas the second suggests that a separate pathway is affected for each of the two phenotypes (106).

4.1.5.2.1 Microtubule stabilisation and maintenance of the primary cilium

Primary cilia act as chemical and mechanical sensors on the surface of cells (176). Those present on renal epithelial cells have a variety of roles including responding to changes in

urine flow, maintaining the nephrons and aiding in development. pVHL binds and stabilises microtubules that are important in cilia maintenance (177).

4.1.5.2.2 Regulation of the extracellular matrix assembly

pVHL has been shown to be essential for the assembly of the extracellular fibronectin matrix (178). pVHL-negative RCC cells show hugely defective extracellular fibronectin matrix assembly, which is restored by reintroduction of wild type pVHL (178).

4.1.5.2.3 Regulation of apoptosis

pVHL has been shown to modulate a number of apoptosis regulators (179). For example it activates NFκB by promoting phosphorylation agonist CARD9 (180) and inactivating p53 via inhibiting MDM2-mediated ubiquitination and nuclear export (181).

4.1.5.2.4 Control of cell senescence

VHL inactivation results in cell senescence via activation of pRB and down regulation of p400 (a SWI/SNF chromatin remodeling protein) (182).

4.1.5.2.5 Transcriptional regulation

An example of HIF-independent VHL transcriptional regulation is that pVHL targets RNA polymerase II for ubiquitination, using the same binding site as HIF (183).

4.1.6 VHL mutation spectrum

Germline *VHL* mutations have been reported in more than 900 families with VHL disease (144;184) which has allowed genotype-phenotype correlations to be made. The first 53 amino acids of $pVHL_{30}$ display poor evolutionary conservation and are not present in $pVHL_{19}$. Only a handful of potentially pathogenic mutations have been reported in this domain (144;184). The majority of patients with truncating mutations or exon deletions have Type 1 VHL disease (Table 4.1) (144). A subset of Type 1 patients have a pathogenic full or partial *VHL* deletion where the deletion encompasses *BRK1,* an adjacent gene. These individuals develop retinal and CNS HABs but have a low-risk of ccRCC. This phenotype has been referred to as Type 1B (139;185-187). The lack of ccRCC is thought to result from the combination of *BRK1* and *VHL* inactivation causing renal cells to be defective in proliferation, polarisation and motility, and therefore inefficient in tumour formation (188). The majority of Type 1 deletions are mediated by Alu repeats which occur with comparatively high frequency in the *VHL* genomic region (139).Type 2 VHL disease is mainly caused by germline missense mutations (84%) (144). The majority of kindreds have RCC, HBB and PHEO and are considered to be Type 2B. Protein conformation analysis has suggested that surface amino acid substitutions yield a higher risk of PHEO than those closer to the protein core (189).

In vitro modelling of *VHL* mutations has provided evidence that the risk of developing HAB or ccRCC is associated with the ability of mutant pVHL to impair HIF activity (121;157;158). Major defects in HIF regulation are associated with Type 1 and Type 2B VHL mutations, whereas Type 2A mutations appear to be less defective. Some Type 2C VHL disease mutations can still downregulate HIF1α (121;157) , which means VHL-related PHEOs are likely to be caused by a HIF-independent mechanism. It has been suggested that compete

loss of pVHL function is lethal/disadvantageous for PHEO precursor cells (106). These correlations are complicated by the high likelihood that genetic modifiers also influence VHL disease expression (190-192).

4.1.7 Testing VHL to date

Diagnostic molecular genetic testing currently comprises sequence analysis of the three *VHL* exons and a small amount of adjacent intronic material, either alone or as part of a panel of genes (193), and dosage analysis of the three exons using MLPA. This methodology is thought to identify the vast majority of *VHL* mutations (>95%). However, there are affected individuals in whom a pathogenic mutation has not been identified, and there are a several possible reasons why mutations in *VHL* could be missed by the routine screening strategy.

Firstly, a number of regions of the VHL gene are not routinely analysed, including the promotor/5' untranslated region (UTR), the 'deep' introns, the 3'UTR and other potential regulatory elements. Pathogenic mutations, which are associated with other conditions, have been identified in these regions of other genes (194), and it therefore seems possible that this could be the case for *VHL* too.

Secondly, it is also conceivable that germline promotor hypermethlylation of *VHL* occurs, as it does in *MLH1* for Lynch syndrome, since this can be one of the 'hits' in sporadic ccRCC cases (195).

Thirdly, whilst dosage analysis will affect deletions and duplications of the gene, it cannot detect structural rearrangements which may disrupt the gene.

Fourthly, mosaic level mutations have previously been identified in VHL probands (196;197). The majority of diagnostic laboratories either use a conventional PCR and Sanger sequencing approach, which is not designed to identify mosaic mutations although they may be observed to varying levels in different laboratories, or NGS workflows designed to confidently identify heterozygous mutations (193).

Finally, the pathogenic mutation may be in another gene. A recent example of this is the identification of *SUFU* as a gene causing Gorlin syndrome, which was thought to be monogenic with mutations in *PTCH1* alone (198).

4.2 Aims

Not all patients with a clinical diagnosis of VHL disease have had their pathogenic *VHL* gene mutation identified. The identification of a mutation is of importance as it allows appropriate management of individuals within a family. The aim of the research reported in this chapter of the thesis was to design an NGS assay in order to comprehensively search for the 'missing' sequence mutations in *VHL* by screening the entire gene region at high depth of coverage in order to be able to identify variants in regions not commonly analysed and/or at mosaic level.

4.3 Materials and methods

4.3.1 DNA Samples

The study was comprised of a development phase, a blind validation phase and a diagnostic phase. Phase I encompassed validating the system using a set of 75 DNAs with known Sanger sequencing and multiplex ligation dependent probe amplification (MLPA) results where a likely pathogenic variant had been identified (Table 4.2).

Table 4.2 Types of Class IV and Class V variants used to validate the assay
Variants classed as IV (likely pathogenic) or V (pathogenic) based on the 'Classification System for Sequence Variants Identified by Genetic Testing' proposed by Plon *et. al.*(199).

Variant type	Class V	Class IV	TOTAL
Large deletions	29	0	29
Indels	15	0	15
Nonsense	3	0	3
Missense	15	11	26
Inframe	0	1	1
Intronic	0	1	1
TOTAL	62	13	75

Phase II comprised confirming the capacity of the system to identify mosaicism. This was achieved by creating admixtures of a common DNA with 4 DNAs at 5 different ratios (50%, 25%. 12.50% 6.25% and 3.13%) in addition to processing those 4 DNAs alone. The variants in each of the DNAs are described in Table 4.3

Table 4.3 Variants included in the mosaicism experiment

Sample	Variant
DNA 1	c.[212_213delCC];[=] p.(Pro71Leufs*60)
DNA 2	c.[238dupA];[=] p.(Ser80Lysfs*52)
DNA 3	c.[407T>C];[=] p.(Phe136Ser)
DNA 4	c.[500G>A];[=] p.(Arg167Gln)
DNA 5	c.[525C>A];[=] p.(Tyr175Ter)

Phase III was designed to search for causative mutations in those DNAs where standard *VHL* analysis had not identified a likely pathogenic mutation. Fifty DNAs were analysed, 41 DNAs where previous Sanger sequencing had identified no variants (excluding known Class I polymorphisms) and an additional 9 DNAs where the only variant identified has been given a Class II - Likely Benign or Class III - Variant of Uncertain Significance status.

4.3.2 Next Generation Sequencing (NGS) assay

NGS analysis was performed on a 15, 956 bp genomic region (ch3: 10, 179, 440-10, 195, 396 hg19) containing the *VHL* gene and its flanking sequences. Peripheral leukocyte DNA was amplified in 3 overlapping fragments of 6, 158 bp, 6, 487 bp and 4, 690 bp respectively. The genomic region was selected to include the entire VHL gene region and flanking DNA. These fragments were amplified using the SequalPrep Long PCR kit (Life Technologies). Amplicons were electrophoresed on a 1% W/V TBE gel for quality control. Then the amplicons for each DNA were pooled in equimolar concentrations, indexed and prepared for sequencing using the Nextera XT kit (Illumina), in accordance with the manufacturer's instructions.

Sequencing was performed using the Illumina MiSeq (Illumina), using 2 x 151 bp reads, again in accordance with the manufacturer's instructions. Twenty four DNAs were analyzed per MiSeq run.

Six NGS runs were performed by the author alone and four with assistance from Shaun Green or Vera Cercqueira on behalf of the WMRGL.

4.3.3 NGS data analysis

Data analysis was performed using NextGENe v2.3 sequence analysis software (SoftGenetics) and a semi-automated in-house bioinformatics pipeline. After removal of poor quality reads (median score threshold <20, maximum number of uncalled bases <2, called base number of each read <100 and/or <3 bases with score of <16), samples were demultiplexed and reads mapped against gene-specific GenBank reference files. Reads were aligned with a matching requirement of >=12 bases with >=85% coverage, with 'detect large indels' and 'rigorous alignment' selected. Data was analysed by 2 methods. 1) The entire ~16kb region was examined for variants in >5% of >30 total reads. 2) The coding region was analysed for variants in > 1% of >30 total reads.

Multiple variants were identified for each DNA. Variants were removed if considered to be a Class I – Benign variant or an artefact. A Class I – Benign variant was defined as a variant present in dbSNP132 and the 1000 genomes project with a minimum allele frequency of greater than 1% in 1 or more of the core populations (African (AFR), American (AMR), Asian (ASN) and European (EUR)). An artefact was defined as any intronic variant that occurred in a homopolymer tract and was present in at least 50% of the samples on the run, or any variant that was called in at least 1 sample but observed in all samples within a run at a broadly similar level. Reads containing likely artefacts were examined and lack of consistency with a known variant was also considered. The Class I variants and artefacts excluded from further analysis are listed in Appendix Table 7.2 and Table 7.3, respectively.

4.3.4 Variant assessment

Following variant identification and confirmation, determination of variant pathogenicity was performed according to previously published best practice guidelines (200). Much of this analysis was performed using Alamut as an interface (Interactive Biosoftware). This involved a) interrogation of locus specific databases, Human Gene Mutation Database and in house databases, b) literature review, c) search of NHLBI Exome Sequencing Project Exome Variant Server and 1000 genomes project databases, d) species conservation and e) *in silico* prediction for the impact of missense and splicing variants, as appropriate. The data was summarized and an informed assessment was made of likely pathogenicity.

4.4 Results

4.4.1 Phase I

The first phase encompassed blind validation of the system using a set of 75 DNAs with known Sanger sequencing and multiplex ligation dependent probe amplification (MLPA) results where a likely pathogenic variant had been identified. Table 4.4 provides a list of the 24 small scale Class VI and V mutations which had previously been identified by Sanger sequencing in 46 patient DNAs. All the sequence changes were identified by the pipeline. Table 4.5 lists all the whole exon deletions that had previously been identified by MLPA. A number of the exon 1 deletions were visible upon gel electrophoresis of the long PCR product. In addition the deleted region showed a drop in coverage when compared visually to the standard coverage profile (Figure 4.3). The remaining deletions were not visible upon gel electrophoresis or upon NGS. Called sequence variants were examined for zygosity, where the examined region was divided into the three amplicons and regions of overlap ignored. Lack of any heterozygous calls was considered indicative of potential hemizygosity for that amplicon.

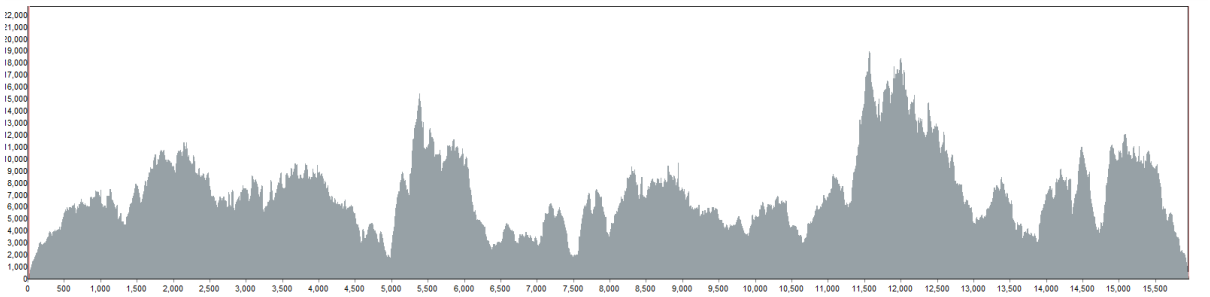
Table 4.4 Small scale *VHL* sequence mutations used for Phase I blind validation of the system

DNA Variant	Protein description	Variant Type	Class	Number of patients
c.[163dupG];[=]	p.(Glu55Glyfs*77)	Indel	V	6
c.[194delC];[=]	p.(Ser65Trpfs*2)	Indel	V	2
c.[227_229delTCT];[=]	p.(Phe76del)	Inframe	IV	1
c.[233A>G];[=]	p.(Asn78Ser)	Missense	V	5
c.[239G>A];[=]	p.(Ser80Asn)	Missense	IV	1
c.[240T>G];[=]	p.(Ser80Arg)	Missense	IV	1
c.[250G>A];[=]	p.(Val84Met)	Missense	IV	3
c.[250G>T];[=]	p.(Val84Leu)	Missense	IV	1
c.[257C>T];[=]	p.(Pro86Leu)	Missense	V	1
c.[262T>G];[=]	p.(Trp88Gly)	Missense	IV	3
c.[292_295delTACC];[=]	p.(Tyr98Glnfs*60)	Indel	V	1
c.[326_340+16del31insA];[=]	p.(Ile109Lysfs*18)	Indel	V	1
c.[332G>A];[=]	p.(Ser111Asn)	Missense	IV	2
c.[337C>T];[=]	p.(Arg113Ter)	Nonsense	V	1
c.[361delG];[=]	p.(Asp121Metfs*38)	Indel	V	1
c.[369delG];[=]	p.(Thr124Hisfs*35)	Indel	V	1
c.[463+3A>T];[=]		Intronic	IV	1
c.[479_480delAG];[=]	p.(Glu160Alafs*13)	Indel	V	1
c.[481C>T];[=]	p.(Arg113Ter)	Nonsense	V	2
c.[486C>G];[=]	p.(Cys162Trp)	Missense	V	1
c.[499C>T];[=]	p.(Arg167Trp)	Missense	V	7
c.[500G>A];[=]	p.(Arg167Gln)	Missense	V	1
c.[509delT];[=]	p.(Val170Alafs*32)	Indel	V	1
c.[606dupA];[=]	p.(Gln203Thrfs*53)	Indel	V	1

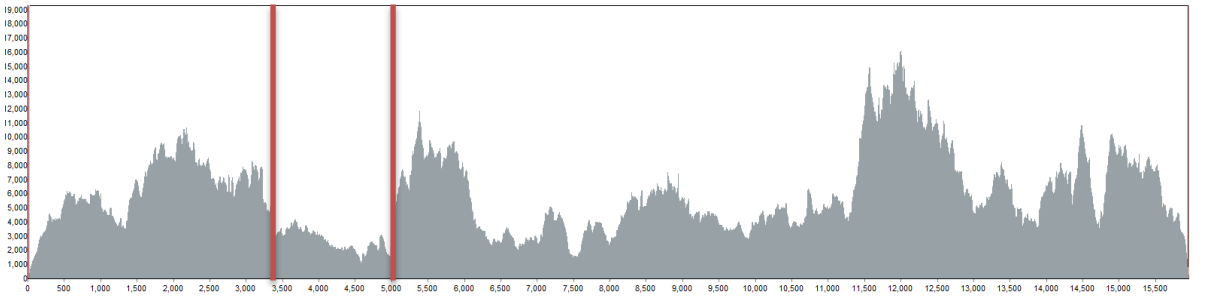
Table 4.5 Results for large deletion mutation included in blind validation
Details of whole exon deletions used for Phase I blind validation of the system, showing if the deletion was visualised upon gel electrophoresis and/or if there were any heterozygous variants (Class I-V) observed in each of the three long PCR amplicons.

DNA Variant	Class	Gel visualised?	Only homozygous variants observed?		
			Amplicon 1	Amplicon 2	Amplicon 3
Exon 1 deletion	V	Yes	Y	Y	Y
Exon 1 deletion	V	Yes	N	N	N
Exon 1 deletion	V	Yes	Y	Y	Y
Exon 1 deletion	V	Yes	Y	Y	Y
Exon 1 deletion	V	Yes	Y	Y	Y
Exon 1 deletion	V	Yes	Y	Y	Y
Exon 1 deletion	V	Yes	N	N	N
Exon 1 deletion	V	Yes	Y	Y	Y
Exon 1-2 deletion	V	No	Y	Y	N
Exon 1-2 deletion	V	No	Y	Y	N
Exon 1-2 deletion	V	No	Y	Y	Y
Exon 1-2 deletion	V	No	Y	Y	Y
Exon 1-2 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 1-3 deletion	V	No	Y	Y	Y
Exon 3 deletion	V	No	N	Y	Y
Exon 3 deletion	V	No	N	N	Y
Exon 3 deletion	V	No	N	N	Y

a)



b)



c)

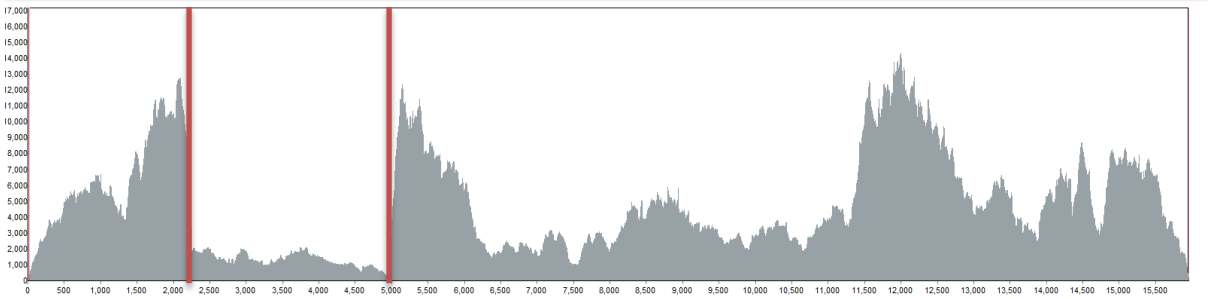


Figure 4.3 Coverage graphs a) Standard coverage profile; b) Example deletion profile, estimated size of deletion ~1, 600 bp; c) second deletion profile, estimated size of deletion ~2, 800 bp.

4.4.2 Phase II

The capacity of the system to identify mosaicism was validated by spiking four DNAs containing a known heterozygous variant (DNA 1-4) with a common DNA, also containing a known heterozygous variant (DNA 5) at 6 different admixtures (Table 4.6).

Table 4.6 DNA admixtures for mosaic validation panel

Low admixture samples		**High admixture sample**	
DNA 1-4 (% of sample)	**Frequency variant allele expected (%)**	**DNA 5 (% of sample)**	**Frequency variant allele expected (%)**
100.00%	50.00%	0.00%	0.00%
50.00%	25.00%	50.00%	25.00%
25.00%	12.50%	75.00%	37.50%
12.50%	6.25%	87.50%	43.75%
6.25%	3.13%	93.75%	46.88%
3.13%	1.56%	96.87%	48.44%

The percentage of reads in which each variant was called at each admixture and the depth of coverage were recorded. All variants at all admixtures were identified (Table 4.7). All variants above 5% were called by the 5% pipeline, with no additional coding variants called. Variants at 1.5% and 3.13% were called using the 1% pipeline. In this pipeline additional coding variants were called at a very low level; however these erroneous variants were present in all DNAs and therefore excluded from analysis as an artefact. The assay was designed to maintain a very high coverage depth in order to increase the probability of identifying variants at a low level. Variants expected to be observed at 1.56% were all identified at an average of 2.0% (Range: 1.50-2.05%) with an average read depth of 2809 (Range: 798-5992).

Table 4.7 Results of mosaicism identification experiment for: a) low level admixture samples and b) high level admixture samples.

a)

Frequency variant allele expected (%)	Mix 1 (DNA1 and 5) c.212_213delCC		Mix 2 (DNA2 and 5) c.238dupA		Mix 3 (DNA3 and 5) c.407T>C		Mix 4 (DNA4 and 5) c.500G>A		Average
	Variant %	Coverage Depth	Variant %	Coverage Depth	Variant %	Coverage Depth	Variant %	Coverage Depth	
50.00%	46.88%	480	43.86%	4099	44.73%	7299	46.40%	11682	45.78%
25.00%	21.42%	985	27.53%	2815	24.79%	5213	24.00%	8390	24.50%
12.50%	10.70%	2598	13.17%	4291	14.07%	6190	12.35%	13789	12.70%
6.25%	4.61%	347	6.27%	3415	7.55%	6307	7.30%	8955	6.53%
3.13%	2.47%	1297	4.38%	3150	4.00%	8043	3.10%	15960	3.50%
1.56%	1.50%	798	2.05%	3169	1.80%	5992	1.90%	1277	2.00%

b)

Frequency variant allele expected (%)	Mix 1 (DNA1 and 5) c.525C>A		Mix 2 (DNA2 and 5) c.525C>A		Mix 3 (DNA3 and 5) c.525C>A		Mix 4 (DNA4 and 5) c.525C>A		Average
	Variant %	Coverage Depth	Variant %	Coverage Depth	Variant %	Coverage Depth	Variant %	Coverage Depth	
0.00%	0.00%	1817	0.00%	15255	0.00%	11136	0.00%	11942	0.00%
25.00%	18.65%	3228	18.24%	9587	19.50%	8788	19.62%	8222	19.21%
37.50%	30.18%	7943	29.81%	13548	29.44%	9864	29.29%	13279	29.94%
43.75%	37.19%	847	36.52%	11087	36.58%	10936	35.12%	8564	36.86%
46.88%	41.24%	3615	39.15%	9587	40.41%	13028	39.37%	14924	40.53%
48.44%	43.54%	2269	42.22%	9244	42.03%	9495	42.56%	1189	43.36%

4.4.3 Phase III

The third phase involved prospective analysis of an additional 50 DNAs, in which either no reported variant or a variant in Class II or Class III had been identified previously (Table 4.8). In this phase all 9 of the Class II/III variants that had been previously identified by Sanger sequencing analysis were confirmed. For five of those DNAs no additional sequence variants were identified, but in four DNAs an additional one or two heterozygous non-coding variants were observed. These variants were classified as II or III and therefore were not confirmed to be pathogenic. Phase III patients were not interrogated for sequence variant zygosity as all had previously been analysed for dosage by MLPA in a diagnostic laboratory setting.

Table 4.8 Class II and III variants, previously identified and confirmed by this assay

Variant call	Protein call	Variant Class
c.[183C>G];[=]	p.(Pro61Pro)	II
c.[242C>T];[=]	p.(Pro81Leu)	III
c.[338G>A];[=]	p.(Arg113Gln)	III
c.[341-21_341-17delAACCT];[=]		III
c.[361G>C];[=]	p.(Asp121His)	III
c.[463+8C>T];[=]		III
c.[464-10G>A];[=]		III
c.[554A>G];[=]	p.(Tyr185Cys)	III
c.[639T>C];[=]	p.(Asp213Asp)	III

Forty one patient DNA samples were examined where no previous variant or mutation had been identified by Sanger sequencing of the exons or MLPA dosage analysis. In 26 DNA samples no sequence variants were identified. In 12 DNAs Class II or III variants were observed, specifically 8 contained a Class III sequence variant and 4 DNAs had 1 or 2 Class II sequence variants (Table 4.10). Finally, 3 DNAs contained a mosaic level Class IV or V mutation (Table 4.11).

Whilst exon level dosage examination was considered outside the remit of the test, variant calls were examined to look for samples showing heterozygous calls that were recurrently outside the normal range. One sample gave multiple calls throughout the analysed region at an approximately 60:40 ratio rather than an approximately 50:50 ratio. This finding could indicate a mosaic full gene dosage change. As with all included samples, it had previously been analysed by MLPA in a diagnostic setting and a normal result recorded. The sample also had one Class III variant, c.[340+203G>A];[=].

Table 4.9 Additional variants identified in samples with previously identified Class II or III variants

Variant call	Protein call	Additional variant identified	Class	Brief reason for classification
c.[338G>A];[=]	p.(Arg113Gln)	c.[464-1212A>G];[=]	II	Observed with 1 x Class V variant in another sample
c.[341-21_341-17delAACCT];[=]		c.[*3482dupA];[=]	II	Observed with 1 x Class V variant in another sample, no splicing alterations predicted
c.[464-10G>A];[=]		c.[-3197_-3195delCTC];[=]	III	Not observed with Class IV or V variant, no additional information
c.[639T>C];[=]	p.(=)	c.[*2511A>G];[=]	III	Not observed with Class IV or V variant, Ensembl VEF predicts it is in a regulatory element.
		c.[*3021T>C];[=]	III	Not observed with Class IV or V variant, no additional information

Table 4.10 Variants identified in samples which did not have a variant identified previously

Variants identified	Additional variant	Reason
1 x Class III non-coding variant	c.[-3933C>T];[=]	Not observed with Class IV or V variant, no additional information
1 x Class III non-coding variant	c.[-3933C>T];[=]	Not observed with Class IV or V variant, no additional information
1 x Class III non-coding variant	c.[-963G>A];[=]	Not observed with Class IV or V variant, no additional information
1 x Class II non-coding variant	c.[340+376C>A];[=]	Observed with 3 x Class V variant in other samples
1 x Class III non-coding variant	c.[340+203G>A];[=]	Not observed with Class IV or V variant, no additional information
2 x Class II non-coding variants	c.[340+280T>G];[=]	Observed with 1 x Class V variant in another sample
	c.[464-1434C>T];[=]	Observed with 1 x Class V variant in another sample
1 x Class III non-coding variant	c.[341-123G>T];[=]	Not observed with Class IV or V variant, no additional information
1 x Class III non-coding variant	c.[464-1530C>T];[=]	Not observed with Class IV or V variant, no additional information
1 x Class II non-coding variant	c.[464-1434C>T];[=]	Observed with 1 x Class V variant in another sample
1 x Class II non-coding variant	c.[*820A>G];[=]	Not observed at high enough frequency in stated population groups but in 2% of GBR alleles (1000 genomes)
1 x Class III non-coding variant	c.[*3082C>T];[=]	Not observed with Class IV or V variant, no additional information
1 x Class III non-coding variant	c.[*3170G>A];[=]	Not observed with Class IV or V variant, no additional information

Table 4.11 Details of mosaic mutations identified and level at which identified

Variant call	Protein call	Class	Replicate 1		Replicate 2	
			Reads containing variant base (%)	Depth	Reads containing variant base (%)	Depth
c.[394C=/394C>T]	p.(Gln132Ter)	V	6.75	10723	6.74	1915
c.[277G=/277G>C]	p.(Gly93Arg)	IV	5.78	4635	6.78	1623
c.[499C=/499C>T]	p.(Arg167Trp)	V	1.45	6563	0.92	3696

A number of clinical subgroups were represented in those patients where no likely pathogenic mutation had been found previously (n=50) (Table 4.12). Of those diagnosed clinically with VHL, three of the 17 patients were found to have likely pathogenic mutations at mosaic level in their lymphocyte DNA (see section 4.6 for case studies). In addition, three had a previously identified Class III variant, specifically one patient had a missense variant and two had an intronic variant where splicing analysis may be of benefit, and in five patients a novel Class III variant was identified (Table 4.13). No variant was identified for the remaining 7 patients.

Table 4.12 Reported phenotype of negative VHL gene screen patients

Reported phenotype	Number of individuals
Clinically diagnosed VHL	17
Single tumour type	22
• HAB only	• 17
• Retinal angioma only	• 2
• ccRCC only	• 2
• PHEO only	• 1
Unknown	11

Table 4.13 Class III variants in clinically diagnosed VHL patients

Variant	Number of times identified in cohort (n=17)	Identified in previous analysis
c.[-3933C>T];[=]	2	No
c.[-963G>A];[=]	1	No
c.[341-21_341-17delAACCT];[=]	1	Yes
c.[361G>C];[=] p.(Asp121His)	1	Yes
c.[463+8C>T];[=]	1	Yes
c.[*820A>G];[=]	1	No
c.[*3082C>T];[=]	1	No

Of those patients with a single tumour type (n=22), which alone is not diagnostic of VHL, 14 had no variant identified, 5 had a Class III variant identified worthy of further analysis and 2 had a Class II variant identified.

Table 4.14 Class II variants in patients referred for analysis with a single tumour type

Variant	Number of times identified in cohort (n=22)	Identified in previous analysis	Tumour type
c.[242C>T];[=] p.(Pro81Leu)	1	Yes	PHEO
c.[338G>A];[=] p.(Arg113Gln)	1	Yes	HAB
c.[340+203G>A];[=]	1	No	HAB
c.[341-123G>T];[=]	1	No	HAB
c.[464-1530C>T];[=]	1	No	HAB

For the remaining 12 patients the clinical details could not be accessed.

4.5 Discussion

Current routine practice for the identification of VHL mutations involves MLPA dosage analysis for large deletions and duplications and sequence analysis, using either Sanger sequencing or next generation sequencing, to examine the coding regions and a limited amount of the flanking intronic sequence (193). Recent publications looking at small numbers of patients have demonstrated the potential for using NGS to look for mosaic level mutations (196;197). A strategy was designed to interrogate VHL for both heterozygous and mosaic mutations in the entire VHL gene region to a 5% variant level. Additionally, the coding region was examined for variants down to a 1% level.

Phase I encompassed a blind validation using 75 DNAs. All 24 of the heterozygous small scale (<40 base pairs) sequence variants were identified correctly, providing a sensitivity of 100% (95% CI: 85.62-100.00%).

Whole exon deletions and duplications were outside the expected detection capability of the assay. However during Phase I those deletions which fell inside an amplicon were identified upon gel electrophoresis and NGS allowed further visualisation of the deletion. Eight DNAs with exon 1 deletions were identified by this strategy. While additional exon level deletions could not be visualised, they could be predicted due to the absence of heterozygous calls within the amplicon, indicating potential hemizygosity.

Phase II confirmed the capacity of the system to identify low level mutations by analysing admixtures of two heterozygous variants to create a range of ‘mosaic level’ variants of different types. All the variants were identified down to the lowest admixture (1.5%), using

the two pipelines. This demonstrated the capability of the assay to identify very low level variants.

The third phase of the study involved the analysis of 50 new DNAs, where a likely pathogenic mutation had not been identified by conventional VHL gene analysis. In three patient samples a mosaic mutation was identified within the coding region of the gene. All of the mosaic variants were identified in patients with clinically diagnosed VHL and the identified variant was considered to be the likely cause of their disease (see section 4.6). It is likely that mosaic whole exon or gene copy number changes are also responsible for some of these cases, and a potential example of this was identified in this dataset, but reliably detecting and confirming these variants remains difficult.

Of the remaining 47 patients, 26 had no variants of interest identified and 21 had at least 1 previously or newly identified variant, which were spread throughout the gene region analysed. Unfortunately none of these variants could be classed as likely to be pathogenic. Of most interest are the 5 variants identified in patients with clinically diagnosed VHL disease, in whom a mutation should be identified. It is noteworthy that one of the variants, c.-3933C>T, was identified in two unrelated patients, but in less than 1% the 1000 genomes populations. Additional work has been performed, and is reported in Chapter 5, to attempt to ascertain the pathogenicity of these variants.

There were 6 patients who were clinically described as having VHL disease but in whom no mutation or variants was identified. There are several possible reasons for this, firstly that there is another rare, as yet undiscovered, gene which causes this phenotype or secondly, and perhaps more likely, that this test is not identifying all the mutations in the *VHL* gene.

For example, neither a mosaic large copy number imbalance nor a mutation which acts at very long range would be detected by this analysis.

4.6 Case reports – mosaic mutations

4.6.1 Case I - c.[394C=/394C>T] p.(Gln132Ter) at approximately 7%

Case I was a 43 year old man with a clinical diagnosis of VHL. He had bilateral ccRCC, cerebellar HAB and a suspected spinal HAB. Standard diagnostic *VHL* sequencing analysis and deletion and duplication analysis did not identify a pathogenic mutation. This methodology identified a mosaic nonsense mutation at approximately 7%, c.[394C=/394C>T] p.(Gln132Ter) in exon 2 of VHL. This mutation would cause the protein to be truncated early and would be likely to be subject to nonsense mediated decay (NMD). It has been reported previously in the literature in patients with Type 1 and Type 2B VHL disease (144) and can be considered as Class V – Pathogenic.

4.6.2 Case II - c.[277G=/277G>C] p.(Gly93Arg) at approximately 6%

Case II is a 37 year old male with a clinical diagnosis of VHL. He had PHEO, four cerebellar HAB and bilateral ccRCC. Standard diagnostic VHL sequencing analysis, and deletion and duplication analysis, did not identify a pathogenic mutation. This methodology identified a mosaic missense mutation at approximately 6%, c.[277G=/277G>C] p.(Gly93Arg). This mutation has been reported in the literature in a patient with Type 1 VHL disease (144). It is a highly conserved amino acid and additional sequence changes that cause missense mutations at the same amino acid have been reported in the literature. The variant has not had any functional work performed and is considered to be Class IV – Likely pathogenic.

4.6.3 Case III – c.[499C=/499C>T] p.(Arg167Trp) at approximately 1%

Case III is a deceased male with a clinical diagnosis of VHL. He had two cerebellar HABs, a PHEO and renal tumour. He had no family history and his two children were unaffected at over 40 years of age. The assay identified a missense mutation at approximately 1%, c.[499C=/499C>T] p.(Arg167Trp). This mutation is at the limits of the assay's detection. This mutation has been reported in the literature on multiple occasions in patients with all subtypes of VHL disease (144). This variant is considered to be Class V – Pathogenic.

4.7 Conclusion

In summary, for this study a novel diagnostic assay was developed and validated. It was effective in identifying mosaic level variants, with identification of a clinically actionable mutation in 18% of those patients with clinically diagnosed VHL disease without a *VHL* mutation identified by previous analysis methods. It also identified a number of Class III variants which warrant further analysis.

Chapter 5 Variant classification

5.1 Introduction

5.1.1 Identification of variants of unknown significance (VUSs)

The identification of variants of unknown significance (VUSs) has been a well-recognised problem for the diagnostic genetics laboratory since the introduction of mutation scanning and continues to be a problem in genes which have already been extensively examined.

Mutations in the *BRCA1* and *BRCA2* genes are a well-known cause of inherited breast and/or ovarian cancer, and provide a good example of the problems diagnostic genetic laboratories encounter. The identification of a pathogenic mutation in *BRCA1* or *BRCA2* in an affected individual allows presymptomatic screening throughout a family with prophylactic surgery, heightened screening and targeted drug therapy available for those carrying the mutation, whilst those who don't carry the mutation can return to the less costly screening offered to the general population. Both of these genes have a huge variety of pathogenic mutations, many of which occur in only one or two families worldwide.

Even though *BRCA1* or *BRCA2* are probably the most commonly sequenced genes in the genome, when a genetic variant is identified it may still be almost impossible to determine if the change is pathogenic (18). Even in a laboratory that has sequenced in excess of 150, 000 individuals for BRCA1 and 2, every week 1% to 2% of patients have a variant identified that has never been found previously, many of which will be VUSs (18). Given that there is no reason to believe that the *BRCA1* and *BRCA2* genes are any more mutable or unstable than any other gene in the genome (18), these data suggest that the problem of classifying variants will increase proportionally with the number of genes analysed.

5.1.2 Variant classification

The diagnostic genetic community has adopted a five class system to categorise genetic variants based on their risk of pathogenicity (see Table 5.1), originally proposed by Plon *et. al.* (199). This classification system only pertains to variants that if classified as pathogenic would have reasonable penetrance; it does not include variants, such as those identified in genome wide association studies, that only confer slight alterations to risk of disease (199). Ideally a likelihood ratio would be calculated empirically from multiple lines of evidence, but it is recognised that this is often not possible and it is still very useful to assign each variant to one of the designated classes (199).

Table 5.1 Classification system for sequence variants identified by genetic testing (Adapted from (199)).

Class	Description	Probability of being Pathogenic	Descriptors used in this thesis
V	Definitely Pathogenic	>0.99	Pathogenic (V)
IV	Likely Pathogenic	0.95–0.99	Likely pathogenic (IV)
III	Uncertain	0.05–0.949	VUS (III)
II	Likely Not Pathogenic or of Little Clinical Significance	0.001–0.049	Likely benign (II)
I	Not Pathogenic or of No Clinical Significance	<0.001	Benign (I)

5.1.3 Methods for classifying a variant

The methods available for categorising the pathogenicity of a mutation can be divided into direct methods and indirect methods (201;202). These methods are discussed below, with specific focus on autosomal dominant cancer predisposition genes, although some of the general theories can be extrapolated to other genes and inheritance patterns.

5.1.3.1 Direct 1: Co-segregation with disease in pedigrees

One of the least complicated forms of genetic evidence for pathogenicity is the co-segregation of a variant with the cancer phenotype in a family (201). A pathogenic variant will segregate with affected individuals, whereas a benign variant will segregate randomly. Likelihood ratios can be established using a method similar to linkage analysis by examining the segregation of the variant rather than linked markers (201). Penetrance, the potential for phenocopies, non-paternity and the possibility of the variant being in linkage disequilibrium with the causative mutation must be taken into account (201). The main advantage of this approach is that it depends only on the availability of DNA samples from members of the family with a variant but, as with linkage analysis, its power depends on the number of informative meioses (201). The disadvantages are that often families with a variant are small and it can also be hard to obtain the required samples (201). Unfortunately, it is rarely possible to categorise variants as pathogenic on the basis of segregation alone (201).

5.1.3.2 Direct 2: Comparison of allele frequency between cases and controls

This simple method involves comparing the variant frequency in series of cases and controls. The main disadvantage is that most VUSs are very rare, so an extremely large sample size with extensive phenotypic information would be required to demonstrate that a variant is pathogenic (201). Additionally, the variant of interest may only occur in a certain geographical region or ethnic group making appropriate samples hard to obtain (201). This approach is more successfully used as a rapid way of determining neutral variants (201). If 200 controls are genotyped and the variant of interest is shown to have a frequency of >1% it is highly unlikely to be a high-risk variant, in a rare autosomal dominant condition (201).

However it is very difficult to determine a true control population for a late-onset cancer predisposition disease.

The task of identifying likely benign variants using control populations has been facilitated by a number of genetic variation datasets, which are available online. The first of these is the 1, 000 genomes project, which aimed 'to characterize over 95% of variants that are in genomic regions accessible to current high-throughput sequencing technologies and that have allele frequency of 1% or higher in each of five major population groups' (203;204). The major population groups included were African, Ad Mixed American, East Asian, European and South Asian, with the data released in a number of phases and a total of 2507 human genomes analysed (205). Interestingly, the Phase I release data showed that almost all variants found at a frequency of 10% or greater were identified in all the populations studied, whereas 17% of variants found at a frequency of 0.5-5% were observed in only a single ancestry group (204). Given the diverse ancestry of the British population, reference populations reflecting multiple ancestral groups are an invaluable tool.

The 1, 000 genomes project was followed by the Exome Sequencing Project (ESP). The ESP data have also been released in a number of phases. Currently, data are available from 6503 individuals who are sub-divided into those of European American origin and those of African American origin (206). The cohort is made up of a number of well phenotyped populations, many of whom have heart, lung or blood disorders (206).

Finally, the most recent variation database, the Exome Aggregation Consortium (ExAC), which incorporates 60,706 unrelated individuals from many different studies, was released in October 2014 (207). Again the data are subdivided by ancestral origin, this time into

seven populations, African, Ad-Mixed American, East Asian, South Asian, Non-Finnish European, Finnish European and Other. ExAC includes 1,851 of the 1, 000 genomes individuals, 3,936 of the ESP individuals and numerous individuals from other studies (207). When using the ExAC data for analysis of variants identified in cancer predisposition genes it is important to note that ExAC includes 7,601 samples from the Cancer Genome Atlas (207) and to consider the incidence of disease compared to the minor allele frequency for the variant.

Together these three projects have created ready-made datasets which can be interrogated when analysing any identified variant. The large numbers of samples genotyped mean that individuals with any given cancer predisposition disease are likely to have been included but as long as this is considered when using the data, they provide an invaluable resource.

5.1.3.3 Direct 3: Severity of personal and family cancer history

One hypothesis is that the fact a variant associated with disease will tend to occur in families with a stronger history of disease can be used to aid pathogenicity assessment (201). This theory is based on the premise that an individual with a pathogenic mutation will tend to have a family history similar to that of individuals with other known pathogenic mutations, while a neutral variant will tend to have a weaker family history comparable to individuals without a causative germline mutation (201). The advantage of this approach is that it only requires genotyping of a single individual per family (201). However, it does require detailed data about all the families on whom mutation testing has been performed, to allow comparison (201). This information is often difficult to obtain and those families

which have been screened will have been chosen for their high-risk pedigrees, causing an ascertainment bias.

5.1.3.4 Direct 4: Co-occurrence of a variant of interest with a pathogenic variant

In dominant disease, the co-occurrence *in trans* of a variant with a known pathogenic mutation can be used to refute pathogenicity. Obviously this is dependent on gene and disease specific information, for example mutations on both alleles of a *BRCA1* gene are thought to be embryonically lethal but mutations on both alleles of a *BRCA2* gene produce a Fanconi anaemia phenotype rather than a breast cancer susceptibility one (201). Thus a variant *in trans* with a clearly pathogenic *BRCA1* mutation implies that the variant is benign but in *BRCA2* the patient should be examined for a Fanconi anaemia phenotype which, when ruled out, implies the variant is benign (201). This is a straightforward method of determining pathogenicity but relies on establishing phase. Phase is most easily determined by analysing the genotype of the parents; however the parents will not always be available to provide a sample. Moreover, this approach can only contribute towards ruling out pathogenicity.

5.1.3.5 Direct 5: De novo appearance of the mutation with sporadic disease

The concurrent *de novo* appearance of a variant in a known disease-associated gene with sporadic disease in a family is good evidence of pathogenicity (63). The labelling of a variant as *de novo* relies on absolute determination that the variant is not present in unaffected parents, and the possibility of the assay not detecting the variant in a parent and non-paternity must both be considered. Additionally it is reported that every individual has an

average of 74 *de novo* single nucleotide variations and whilst these are generally more likely to be deleterious, as a result of the lack of evolutionary selection (208), it should considered that a neutral change could have occurred.

5.1.3.6 Direct 6: Tumour pathology methods

Tumour analysis is only possible and of use when examining tumour suppressor genes (TSGs). TSG pathogenicity relies on Knudson's two hit hypothesis and therefore loss of function of the first allele in the germline and the second allele in the tumour (209). Loss of function can occur by mutation, loss of heterozygosity (caused by a deletion or mitotic recombination) or hyper-methylation of the promoter of the second allele of the gene (209). Various methods are available to examine the loss of function of a gene. One of the simplest, if applicable, is immunohistochemistry (IHC), used to detect loss of the associated protein. Loss of heterozygosity (LOH) can also be examined, by looking directly at tumour DNA to establish if the variant is found in the homozygous state, which would indicate deletion of the other allele.

IHC is not always appropriate for variant analysis. For example, loss of function of the VHL is an early event in almost all ccRCC, independent of whether they have a germline mutation (210). Therefore IHC would not be a beneficial. However immunohistochemistry for SDHA and SDHB have been shown to be useful in identifying tumours which are likely to have pathogenic germline mutations in the succinate dehydrogenase genes (73;211). It has been shown that mutations in *SDHB*, *SDHC* and *SDHD* will cause loss of staining for SDHB (212) and mutations in *SDHA* will cause loss of staining in SDHA and SDHB (73). However, it should be noted that not all tumours that show loss of staining will have mutations identified. In

recent studies, between 6% and 16% of tumours without an identified mutation had loss of SDHB staining (211-213).

Tumour analysis is confounded by the difficulty of obtaining samples with good quality DNA. Pathology laboratories routinely formalin fix and paraffin embed tumour samples because this is the best compromise between cost, practicality and the morphological fixation of the tissue (214). Unfortunately, formalin causes cross-linking between DNA stands and degradation leading to low molecular weight DNA (214). This makes the analysis of DNA extracted from FFPE tumour samples difficult to perform.

5.1.3.7 Indirect 1: Assessment of a variant's effect on splicing

Evidence suggests that 15 – 50% of VUSs could influence splicing, including missense, synonymous and intronic variants (215). Intronic changes that sit within the invariant splice sites (AG/GU) almost always alter splicing and in doing so are generally pathogenic (209). The consensus splice site sequences are detailed in Figure 5.1. The prediction of the possible splicing-related effects of a VUS located within the variable bases within the consensus sequences can be difficult and the predictions for those VUSs at a deep intronic position or within an exon are highly complex. Furthermore, exonic VUSs can induce splicing defects instead of, or in addition to, producing functionally relevant protein changes (215). Splicing can be examined *in silico* but data should always be confirmed with an *in vitro* assay.

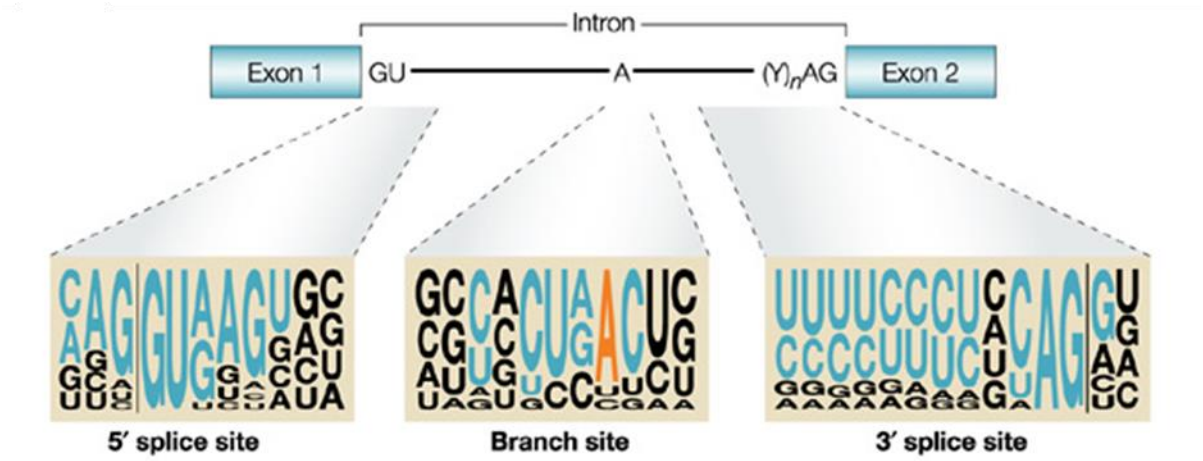


Figure 5.1 Splicing consensus sequences
The three consensus sequences involved in splicing machinery recognition, the 5' splice site (donor site) is 11 bp and comprises the last 3 bp of the exon and first 8 bp of the intron, the branch site is 11 bp and the 3'splice site (acceptor site) is 14 bp and comprises the last 12 bp of the intron and the first 2 bp of the exon. (Adapted from (216)).

5.1.3.7.1 In silico assessment of splicing

In silico splicing assessment can be used to predict pathogenicity. There are numerous programs available (217), all using two classes of algorithm: 1) those that predict splice donor and acceptor sites; 2) those that predict possible exonic or intronic splice enhancer (ESE/ISE) and exonic or intronic splice silencer (ESS/ISS) sites.

The first class of program predicts if a variant disrupts or weakens a true splice site or creates a cryptic splice site in an intron or exon. These predictions tend to be more accurate than those concerning ESE/ISE/ESS/ISS sites, which can be unclear as to the level of effect any disruption may have. ESE/ISE/ESS/ISS predictors search for the many sequences that are potential binding sites of splicing factors, but will not determine which sites are actually used under physiological conditions for a particular gene (218).

There is little published data regarding the validation of splicing prediction programs for diagnostic use (219). In 2012, Houdayer *et. al.* conducted validation upon 327 *BRCA1* and *BRCA2* variants which had undergone both *in silico* and *in vitro* splicing analysis. They used 6 programs, splice site prediction by neural network (NNSplice), splice site finder (SSF), MaxEntScan (MES), human splicing finder (HSF), ESE finder and relative enhancer and silencer classification by unanimous enrichment (RESCUE-ESE) (219). Their findings confirmed that mutation at the invariant splice sites (AG/GU) always affected splicing but that a local cryptic splice site should be considered which may lead to a splicing event other than exon skipping (219). They found that, for the variants examined in *BRCA1* and *BRCA2* within the 5' and 3' splicing consensus sequences, a pathway that combined MES and SSF gave 96% sensitivity and 83% specificity (see Figure 5.1) (219). The pathway used is shown in Figure 5.2. Their study also confirmed the known poor specificity of ESE prediction and branch point prediction (219).

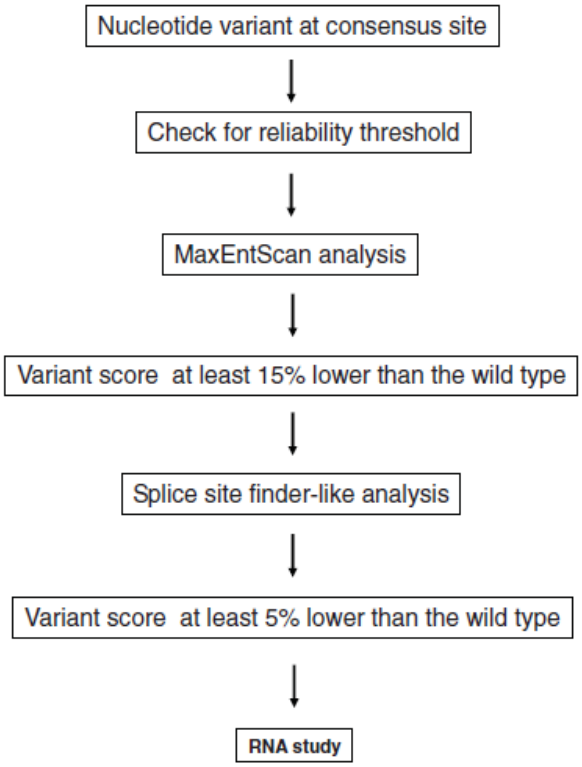


Figure 5.2 Houdayer's recommended analysis pipeline for analysis of variants in the 5' or 3' splice site consensus sequences
Firstly the location of the variant nucleotide should be confirmed to be within the 11 bp the 5' splice site (donor site) or the 14 bp 3'splice site (acceptor site). A reliability threshold should be obtained by defining the mean score and SD for all 5'and 3' splice sites for the given gene and setting the threshold at the mean-2SD (excluding those wild type sites that don't meet this criterion from the pathway). The variant is then considered to require RNA study if the MES score decreases by at least 15% and the SSF score by at least 5%. (Adapted from (219)).

5.1.3.7.2 In vitro assessment of splicing

RT-PCR analysis of patient RNA is the most direct method for assessing splicing aberrations. However its use is hampered by the requirement to obtain RNA from the patient, because RNA is not routinely extracted from blood at the same time as DNA due to cost and, even when it is, the gene of interest may not be expressed in blood (218). Additionally, even if RNA from an expressing tissue can be extracted, analysis is complicated by normal splice isoforms and nonsense mediated decay (NMD) leading to degradation of abnormal products (218). Methods have been created to circumvent some of these problems, including the use

of minigenes (220) and extraction of RNA from lymphoblastoid cell lines treated with puromycin that have been cultured for each patient (221). Unfortunately, not all diagnostic laboratories have access to the cell culture facilities which are required for these two methodologies.

5.1.3.7.3 Indirect 2: Assessment of missense variants

Missense VUSs can be particularly problematic to assess given that they can affect splicing and/or the function of the protein. This section focuses on *in silico* and experimental methods of looking at altered protein function.

5.1.3.7.3.1 In silico assessment of altered function

It is very difficult to design an *in silico* method which is applicable to all missense (MS) variants and genes given that the mechanism by which a MS variant can be pathogenic is variable and genes have numerous functions. Very broadly, the available tools can be subdivided into three groups: 1) Sequence and evolutionary conservation-based methods; 2) Protein sequence and structure-based methods; 3) Supervised learning methods (222). In addition there are some meta-analysis tools which incorporate several tools to make consensus prediction.

Sequence and evolutionary conservation-based methods use the premise that a disease causing variant is more likely to occur at a position that has been conserved during evolution if it is essential to the function and/or structure of the protein (222). Evolutionary conservation can be assessed by aligning homologous genes from multiple species and looking to see what degree of divergence has occurred between them (222). In addition, the physiochemical difference between the wild type and variant amino acid can be calculated.

The quality of the sequence alignment and the number of species included can both directly impact the results produced from these programs (222). There are numerous tools available which use these methods and they include SIFT (223), Align-GVGD (224), Mutation assessor (225), PANTHER (226) and PROVEAN (227).

Protein sequence and structure-based methods use information on the normal proteins structure and sequence to predict the consequence of an amino acid substitution (222). Tools which use this method often take into account how the variant amino acid it likely to change the structure of the protein and/or if it sits within a known domain e.g. a binding domain (222). There are a limited number of programs of this type available and the most commonly used is PolyPhen-2, which has two different algorithms HumDiv and HumVar (228).

The final set of tools use supervised-learning methods. There are various supervised learning algorithms which include neural networks, support vector machines, random forests and naive Bayes classifiers (222). These programs may also use sequence and evolutionary conservation and/or protein sequence and structure but rather than using them directly they are trained with variants of known pathogenicity. Neural networks and support vector machines use two datasets, one comprising known disease mutations and one comprising known benign variants and the algorithm learns to discriminate between the two (222). Random forest methods work by combining the predictions of different methods associated with a pathogenic mutation and when an unknown variant is submitted it assesses the same methods to see if it matches a known pathogenic change and makes a prediction (222). However, all supervised-learning methods rely on the quality of the

datasets used to train them and the number and diversity of the variants included. There are numerous tools of this type including, I-Mutant (229), MutationTaster (230), MutPred (231), nsSNPAnalyzer (232), PON-2P (233), PhD-SNP (229), SNAP (234) and SNPs&GO (235).

There are also tools which perform meta-analyses of a number of these programs. They include MetaSNP, which incorporates PhD-SNP, SIFT and SNAP (236), PredictSNP which incorporates MAPP, nsSNPAnalyzer, PANTHER, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT and SNAP (237), and CONDEL, which incorporates Log R Pfam E-value, MAPP, Mutation Assessor, Polyphen2 and SIFT (238).

A number of studies have looked at the performance of a subset of these prediction tools, however comparatively few are independent. One of the largest studies was performed by Thusberg *et. al.* who evaluated the performance of nine prediction algorithms (MutPred, nsSNPAnalyzer, Panther, PhD-SNP, PolyPhen, PolyPhen2, SIFT, SNAP, and SNPs&GO) using more than 40, 000 variants which had previously been classified into pathogenic and neutral (239). Pathogenic missense variants were identified using the Phencode database, registries in IDbases and 18 locus specific databases (LSDBs) (239). Neutral variants were taken from the dbSNP database and had to have an allele frequency greater than 0.01 and a chromosome count of greater than 49 (239). The group performed statistical analysis on the findings and concluded that SNPs&GO and MutPred were the most reliable predictors in their dataset, because these methods had high accuracy, precision, specificity, sensitivity and negative predictive value (239). However, no single program was rated best by all statistical methods used, which indicates the complexity of attempting to predict the pathogenicity of variants computationally (239). Inevitably, the performance of a program

will vary depending on the gene analysed, with one being more applicable to a specific structure and/or function of the protein than another. While the predictions made by these programs can be used to suggest the likely pathogenicity of a variant, they should never be used alone to assign pathogenicity.

5.1.3.7.3.2 *In vitro* assessment of altered function

Functional assay development is notoriously difficult and, when used for the purpose of assigning or rejecting pathogenicity, the assay must be very well validated. Assays should be designed with the specific gene in mind and even then are very unlikely to encompass all the functions of a multifunctional protein, such as pVHL. Some of the confounding factors in functional analysis are summarised in Figure 5.3, which highlights the ease with which an assay can be misinterpreted. Proving that a variant has no functional effect is very difficult and ideally requires multiple lines of evidence, as a negative result only refers to the specific function analysed (240).

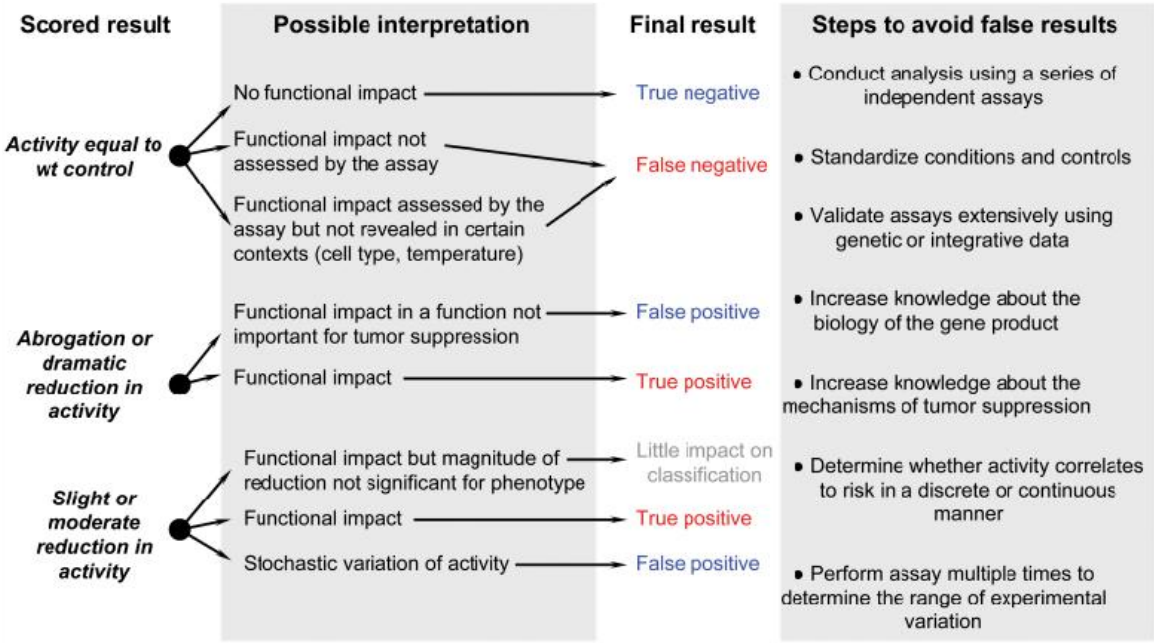


Figure 5.3 Interpretation of functional assays, the pitfalls and steps to avoid false results (Adapted from (240)).

5.2 Aims

The aim of this stage of the study was to undertake further assessment of variants identified during the course of the gene screening, reported in chapters 3 and 4. Having provided a comprehensive list of the ways in which a variant can be assessed, this chapter now goes on to set out the findings from performing those analyses which it was possible to undertake within the laboratory setting and includes:

1) Compilation of all variants identified;
2) Interrogation of currently available variation databases for all Class II and III variants identified;
3) In silico prediction of:
    a. splicing aberrations for all Class II and III variants identified;
    b. protein function for all Class II and III missense variants identified;
4) In vitro analysis of predicted splicing aberrations, where possible;
5) Tumour analysis for loss of protein function, where possible;
6) In vitro functional assessment of protein stability for VHL variants;
7) Compilation of all data obtained to allow reclassification of variants observed.

5.3 Materials and methods

5.3.1 Compilation of all variants identified during the project

All the variants identified during the course of gene screens were compiled in a table.

5.3.2 Variation database interrogation – updated July 2015

Three online variation databases were interrogated for the variants defined at Class II or III during the gene screening period of this research.

- 1000 genomes (super populations: African (AFR); Ad-Mixed American (AMR); East Asian (EAS); European (EUR) and South Asian (SAS)). The discrete populations within each super population were not examined individually due to their smaller sample size (204);
- Exome sequencing project (ESP) (populations: European Americans (EA) and African Americans (AA)) (206) ;
- Exome Aggregation Consortium (ExAC) (populations: African (AFR); Ad-Mixed American (AMR); East Asian (EAS); South Asian (SAS); Non-Finnish European (NFE); Finnish European (FIN) and Other (OTH)) – it should be noted that ExAC incorporates the data from 1000 genomes and ESP and should therefore not be considered an independent source. It also contains samples from the Cancer Genome Atlas. (207).

Once the data were collated those variants that occurred in any population at ≥1% were examined for with the aim of downgrading their pathogenicity class, if appropriate.

5.3.3 In silico variant analysis

5.3.3.1 In silico splicing analysis

In silico splicing analysis was performed with five different algorithms, using Alamut as an interface. The algorithms incorporated were:

- SpliceSiteFinder-like (SSF), which gives a score of 0-100 (241);
- MaxEntScan (MES), which gives a score between 0-12 (242);
- NNSplice (NNS), which gives a score between 0-1 (243);
- GeneSplicer (GS), which gives a score between 0-15 (244);
- Human Splicing Finder (HSF), which gives a score between 0-100 (245).

The results were considered in two ways. Firstly, the pathway established by Houdayer *et.al.* was used to interrogate variants within the 5' and 3'consensus sequences (see Figure 5.2) (219).

Secondly all variants were interrogated with more permissive criteria. It was recorded that splicing may be affected if 1) any true donor or acceptor site was predicted to have a greater than 10% decrease in score by more than one program; 2) any cryptic donor or acceptor site was predicted to have a greater than 10% increase in score by more than one program.

Those samples in which the variant was identified to potentially affect splicing were further examined. For variants that were predicted to abolish a true splice site, the local sequence predictions were examined for cryptic splice sites to allow anticipation of the alternative splicing that may occur. For the variants that were predicted to activate a cryptic splice site, the scores were compared to the average scores less two standard deviations for the actual

donors or acceptors for that gene, as appropriate, to allow prediction of the likelihood of the cryptic splice site's use. Additionally, the splicing scores in the sequence local to the variant base were considered.

5.3.3.2 In silico missense variant analysis

Seventeen *in silico* missense mutation prediction tools, including three meta-servers were used to analyse the variants identified (see Table 5.2). Three programs are routinely used by the diagnostic genetics community in the UK as a result of their integration into Alamut, which is reference software for human variation interpretation used by many diagnostic laboratories. These three programs are AlignGVGD, SIFT and PolyPhen2.

The programs were divided into four categories based on the method used for their prediction: 1) Sequence and evolutionary conservation; 2) Protein sequence and structure; 3) Supervised learning; 4) Meta-analysis. The results of running the programs in each category were tabulated and the concordance (100%) between the predictions of the programs within each category was examined. If there were more than 3 programs in a category 75% concordance was also examined. If the same prediction was obtained for ≥75% of programs in a category that result was taken as the prediction for that category. In addition to the concordance within a category of programs, the concordance between the categories of program was examined.

Table 5.2 In silico missense tools used for analysis of identified Class II and III variants

Name	Method type	Uniform Resource Locator (URL)
Align GVGD (224)	Sequence and evolutionary conservation	http://agvgd.iarc.fr/agvgd_input.php
CONDEL (238)	Meta-predictor	http://bg.upf.edu/condel
I-Mutant (229)	Supervised-learning	http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi
MetaSNP (236)	Meta-predictor	http://snps.biofold.org/meta-snp/
Mutation Assessor (231)	Sequence and evolutionary conservation	http://mutationassessor.org/
Mutation Taster (225)	Supervised learning	http://www.mutationtaster.org/
MutPred (230)	Supervised learning	http://mutpred.mutdb.org/
nsSNPAnalyzer (232)	Supervised learning	http://snpanalyzer.uthsc.edu/
PANTHER (226)	Sequence and evolutionary conservation	http://pantherdb.org/tools/csnpScoreForm.jsp?
PhD-SNP (229)	Supervised-learning	http://snps.path.uab.edu/phd-snp/phd-snp.html
Polyphen-2 (228)	Protein sequence and structure	http://genetics.bwh.harvard.edu/pph2/
PON-P2 (233)	Supervised-learning	http://structure.bmc.lu.se/PON-P2/
PredictSNP (237)	Meta-predictor	http://loschmidt.chemi.muni.cz/predictsnp/
PROVEAN (227)	Sequence and evolutionary conservation	http://provean.jcvi.org/protein_batch_submit.php?species=human
SIFT (223)	Sequence and evolutionary conservation	http://sift.jcvi.org/
SNAP (234)	Supervised-learning	http://www.bio-sof.com/snap
SNPs&GO (235)	Supervised-learning	http://snps-and-go.biocomp.unibo.it/snps-and-go/

5.3.4 RNA analysis

5.3.4.1 PCR of cDNA for analysis of VHL gene splicing

PCR of cDNA was performed with primers designed using the Primer3 software (62). Primers then had a common forward or reverse M13 sequence added as appropriate to allow the use of a common forward or reverse sequencing primer.

Table 5.3 *VHL* cDNA amplification primers

Primer name	Primer sequence (M13 tag underlined)
VHL_cDNA_F_M13	TGTAAAACGACGGCCAGTGAGTACGGCCCTGAAGAAGA
VHL_cDNA_R_M13	CAGGAAACAGCTATGACCTCAATCTCCCATCCGTTGAT
VHL_full_cDNA_F_M13	TGTAAAACGACGGCCAGTATCCACAGCTACCGAGGTCAC
VHL_Δ2_cDNA_F_M13	TGTAAAACGACGGCCAGTCTACCGAGTGTATACTCTGAAAGAGC
VHL_FullΔ2_cDNA_R_M13	CAGGAAACAGCTATGACCCCGCTACGGATGTAGAATGG

Table 5.4 VHL cDNA PCR master mix reagents and volumes

Component	Volume per reaction (µl)
Megamix-W (Microzone)	22
Forward primer (20nM; Sigma)	1
Reverse primer (20nM; Sigma)	1
cDNA (20ng/µl)	1
TOTAL	25

Table 5.5 Thermal cycling conditions for VHL cDNA PCR

Step	Temperature	Time	Cycles
1 (Initial denaturation)	95 °C	5 minutes	1
2 (Denaturation)	95 °C	1 minute	30
3 (Primer annealing)	60 °C	1 minute	
4 (Extension)	72 °C	1 minute	
5 (Final extension)	72 °C	10 minutes	1

5.3.4.2 VHL cDNA PCR agarose gel electrophoresis and imaging

Amplified PCR products were resolved on a 2% (w/v) agarose gel. The gel was made by dissolving 4 g of agarose (Sigma) in 200 ml of 1 x TBE buffer (89 mM Tris, 89 mM Boric Acid;

2 mM EDTA; pH8.3; Severn Biotechnology) by heating in a microwave at 400 W, until the agarose had fully dissolved. Once cooled, 10 ng/ml ethidium bromide was added. Gels were cast in a 20 cm x 30 cm tray, a comb added and left to set. Gel electrophoresis was carried out in a horizontal tank containing 1 x TBE buffer. 5ul of PCR product was mixed with 5ul of 2x loading buffer (10 mg Orange G (Sigma), 1.5 ml Glycerol (Sigma) with ddH_2O added to a final volume of 25 ml) and run on the gel at 100 V for approximately 1 hour and 30 minutes. To approximately determine the size of amplified products, Hyperladder II (Bioline) was also loaded. Gels were viewed using ultraviolet transilluminator (254nm wavelength) and images captured using an Ingenius Syngene Bioimaging machine.

5.3.4.3 Digital Droplet PCR using the Biorad QX200

The PCR reaction components were mixed as per Table 5.7. Two mixes were made for each patient, one using the forward primer designed to amplify the full cDNA of *VHL* and one using the forward primer designed to amplify the isoform lacking exon 2 (Table 5.6). Each mix was added into a well of a droplet generator cartridge and loaded onto the QX100 Droplet Generator (Bio-Rad), which compartmentalised each mix into approximately 20, 000 oil droplets. The mixes were then amplified by thermal cycling, as per Table 5.8. Following amplification, the QX100 Droplet Reader (Bio-Rad) digitally enumerated the number of positive and negative droplets in each mix. The QuantaSoft program (Bio-Rad) was used to analyse the samples, using the total number of droplets, with and without probe signal, to calculate the number of amplicons as copies/µl. The average relative isoform abundance for each patient or control sample was calculated by dividing the average number of copies/µl of the VHL isoform lacking exon 2 by the average number of copies/µl of the full VHL mRNA transcript.

Table 5.6 VHL Droplet PCR Primer and Probe sequences

Primer name	Primer sequence
VHL_full_cDNA_F_RQ	TGTTGACGGACAGCCTATTT
VHL_Δ2_cDNA_F_RQ	CATCCACAGCTACCGAGTGT
VHL_FullΔ2_cDNA_R_RQ	GTAGAGCGACCTGACGATGT
VHL TaqMan Hydrolysis Probe	VIC-GCCTCCAGGTTGTCCGGAGC-MBG

Table 5.7 VHL Droplet cDNA PCR master mix reagents and volumes

Component	Volume per reaction (µl)
2 x Supermix (Biorad)	12.5
VHL Forward primer (10 µM; Sigma)	2.25
VHL Reverse primer (10 µM; Sigma)	2.25
VHL probe (10 µM; Sigma)	0.75
ddH_2O	2.25
cDNA (20 ng/µl)	5
TOTAL	25

Table 5.8 Thermal cycling conditions for VHL Droplet cDNA PCR

Step	Temperature	Time	Cycles
1 (Initial denaturation)	95 ^{o}C	10 minutes	1
2 (Denaturation)	94 ^{o}C	30 seconds	40
3 (Primer annealing/Extension)	65 ^{o}C	1 minute	
4 (Enzyme deactivation)	98 ^{o}C	10 minutes	1

NB – Slow ramping speed between temperatures - 2 ^{o}C/minute

5.3.5 FFPE Tumour analysis

5.3.5.1 Tumour DNA PCR

PCR of DNA extracted from FFPE tumour tissue was performed with primers designed using the Primer3 software (62). Care was taken to keep amplicons short as FFPE DNA can be degraded. Primers then had a common forward or reverse M13 sequence added as appropriate to allow the use of a common forward or reverse sequencing primer. PCR was

carried out using a hot-start PCR enzyme with a high cycle number thermal profile to give increased yield.

Table 5.9 FFPE DNA amplification primers
The M13 forward and reverse sequences are underlined to distinguish from the gene specific sequence.

Primer name	Primer sequence
SDHB_ex3_M13_FFPE_F	TGTAAAACGACGGCCAGTGGATAAGCTAATACATCCAGG
SDHB_ex3_M13_FFPE_R	CAGGAAACAGCTATGACCTTAGGTTGCACAGCAAGTTC
SDHB_ex4_M13_FFPE_F	TGTAAAACGACGGCCAGTCAGCAAGGAGGATCCAGAAG
SDHB_ex4_M13_FFPE_R	CAGGAAACAGCTATGACCACAAATCCTGCCCTGAAAAA
SDHB_ex5_M13_FFPE_F	TGTAAAACGACGGCCAGTTGATGATGGAATCTGATCCT
SDHB_ex5_M13_FFPE_R	CAGGAAACAGCTATGACCCAGATTGAAACAATAAATAGGGA
SDHB_ex6_M13_FFPE_F	TGTAAAACGACGGCCAGTCAGCTGAAGACATAGCAGAG
SDHB_ex6_M13_FFPE_R	CAGGAAACAGCTATGACCCAGCAATCTATTGTCCTCTTG
SDHB_ex8_M13_FFPE_F	TGTAAAACGACGGCCAGTTTTCCCTTTCAGTTTCAGTT
SDHB_ex8_M13_FFPE_R	CAGGAAACAGCTATGACCCTGTATTCATGGAAAACCAAG
SDHC_ex3_M13_FFPE_F	TGTAAAACGACGGCCAGTCGTTATGCAAAATATTAAACCA
SDHC_ex3_M13_FFPE_R	CAGGAAACAGCTATGACCCAACCTTCAGAACTTTCACC
SDHC_ex5_M13_FFPE_F	TGTAAAACGACGGCCAGTTGTAACTTATGAGCAGCTGTG
SDHC_ex5_M13_FFPE_R	CAGGAAACAGCTATGACCTTCACAGAGAAAATGTGCAA
SDHD_ex1_M13_FFPE_F	TGTAAAACGACGGCCAGTGTTGGTGGATGACCTTGA
SDHD_ex1_M13_FFPE_R	CAGGAAACAGCTATGACCGATGAGTCCTCACTTCCATC
SDHD_ex2_M13_FFPE_F	TGTAAAACGACGGCCAGTTGATGTTATGATTTTTTCTTTTTCT
SDHD_ex2_M13_FFPE_R	CAGGAAACAGCTATGACCCAATTCTTCAAAGTATGAAGTCA

Table 5.10 PCR master mix reagents and volumes from DNA extracted from FFPE tumour

Component	Volume per reaction (µl)
GeneAmp 10X PCR Gold Buffer (ThermoFisher)	2.5
$MgCl_2$ (25mM; ThermoFisher)	2.5
dNTPs (2mM; ThermoFisher)	2.5
Forward primer (20nM; Sigma)	1
Reverse primer (20nM; Sigma)	1
AmpliTaq Gold® polymerase (5U/µL; ThermoFisher)	0.15
DNA (concentration as at extraction)	1
ddH_2O	10.35
TOTAL	25

Table 5.11 Thermal cycling conditions for DNA extracted from FFPE tumour

Step	Temperature	Time	Cycles
1 (Initial denaturation)	95 ^{o}C	15 minutes	1
2 (Denaturation)	95 ^{o}C	1 minute	40
3 (Primer annealing)	55 ^{o}C	1 minute	
4 (Extension)	72 ^{o}C	1 minute	
5 (Final extension)	72 ^{o}C	10 minutes	1

5.3.6 Bacterial Assays

5.3.6.1 Vectors

5.3.6.1.1 pcDNA3.1

The open reading frame of VHL had previously been cloned into pcDNA3.1 by Prof E Maher's group. The vector carries an ampicillin selection marker.

5.3.6.1.2 pIRES2-AcGFP1

The pIRES2-AcGFP1 vector (Clontech; see Figure 5.4) is bicistronic and allows the simultaneous expression of green fluorescent protein (AcGFP1) and a cloned protein of interest, which in turn allows determination of the level of the cloned protein relative to the amount of AcGFP. The vector contains an internal ribosome entry site (IRES2) of the encephalomyocarditis virus. When expressed in cells, the protein of interest is translated in a CAP-dependent manner and AcGFP1 is translated in an IRES dependent manner. AcGFP1 is a green fluorescence protein from *Aequore coerulescens* that can be detected by a specific monoclonal antibody. The vector carries a kanamycin selection marker.

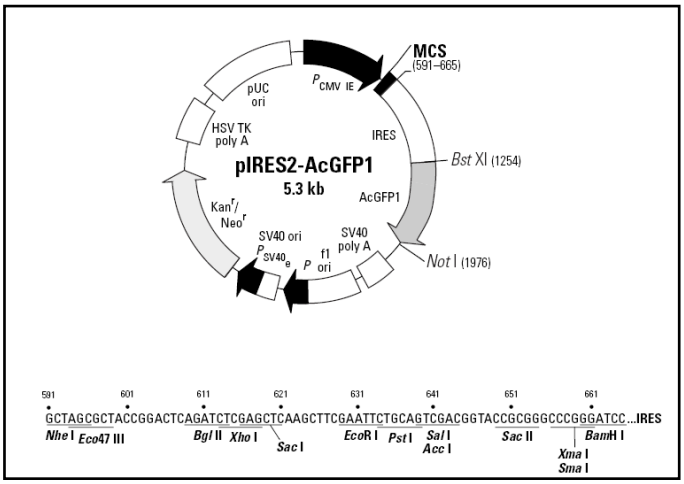


Figure 5.4 pIRES-AcGFP1 vector information and multiple cloning site sequence

5.3.6.2 Bacterial transformation

A general transformation protocol was used for introducing vectors into competent *E.coli* cells, increasing vector quantities and isolating ligated vectors. 5 µl of vector was mixed with 50 µl gold competent *E. coli* cells (Bioline) and incubated for 20 minutes on ice. The tube was then heat shocked for 45 seconds at 42 ^{o}C and returned to ice. 500 µl of SOC medium (Invitrogen) was added to the mix, which was incubated at 37 ^{o}C, shaking at 220RPM for 1 hour. 150 µl of the incubated cells were then plated onto pre-made agar plates containing the required selective antibiotic. The plates were incubated at 37 ^{o}C for 16 hours to allow transformed cells to grow.

5.3.6.3 Plasmid preparation

Plasmid preparation was used to extract and purify plasmid DNA from bacterial cells. It comprised the initial growth of the bacterial culture, lysis of the bacteria and purification of plasmid DNA.

For growth of bacterial cultures, a single colony was picked and transferred into 5 ml of LB broth containing the required selective antibiotic. The cultures were left to propagate overnight at 37 ^{o}C in a shaking incubator (220 rpm).

Two commercial kits were used for bacterial lysis and plasmid DNA purification, depending on the plasmid yield required. A Miniprep kit (Qiagen) was used for yields of 20 – 30 µg of DNA and a Maxiprep kit (Qiagen) was used for yields of 500 – 850 µg. The kits were used as per the manufacturer's instructions, which are described in brief below.

5.3.6.3.1 Miniprep Kit

The bacterial cells were pelleted by centrifugation at 6, 000 x g for 15 minutes at 4 ^{o}C. The cell pellet was re-suspended in 250 µl of buffer P1, containing RNase A. The cells were lysed by addition of 250 µl of buffer P2. This reaction was neutralized by addition of 350 µl of buffer N3. Cells were then centrifuged for 10 minutes at 15,600 x g to pellet the cell debris. The supernatant which contained the plasmid was transferred to a QIAprep spin column and centrifuged at 15600 x g for 30-60 seconds. The flow though was discarded. The DNA bound to the column filter was washed firstly with 500 µl of buffer PB and centrifuged at 15,600 x g for 30-60 seconds and the flow though discarded, followed by repeat centrifugation at 15,600 x g for 60 seconds and the flow though discarded. Secondly the column was washed with 750 µl of buffer PE at 15,600 x g for 30-60 s and then again for 60 s and the flow thorough discarded both times. The column was then transferred to sit on a clean tube and 50 µl of nuclease-free water added. After 1 minute of incubation, the DNA was eluted by centrifugation at 15600 x g for 1 minute.

5.3.6.3.2 Maxiprep Kit

The bacterial cells were pelleted by centrifugation at 6000 x g for 15 minutes at 4 ^{o}C. The cell pellet was resuspended in 10 ml of buffer P1, containing RNase A. The cells were lysed by addition of 10 ml of buffer P2, 4-6 inversions of the tube and incubation at room temperature for 5 minutes. Precipitation of genomic DNA, cell debris and proteins was achieved by addition of 10 ml of buffer P3, 4-6 tube inversions and incubation on ice for 20 minutes. The tube was centrifuged at 20, 000 x g for 30 minutes at 4 ^{o}C to pellet the precipitate. The supernatant was removed and centrifuged again at 20, 000 x g for 15 minutes. The supernatant was removed again. A QIAGEN-tip 500 was equilibrated by

addition of 10 ml of buffer QBT and allowed to empty by gravity flow. The supernatant was added to the tip. The tip was then washed twice with 30 ml of QC buffer which washed the DNA bound to the resin in the tip. The DNA was eluted from the resin with 15ml of buffer QF. The DNA was precipitated by addition of 10.5ml of isopropanol, mixed and centrifuged at 15, 000 x g for 30 minutes at 4 ^{o}C. The pellet was washed twice with 5 ml 70% (v/v) ethanol and centrifuged at 15, 000 x g for 10 minutes. The pellet was left to dry for 10 minutes at room temperature and then eluted in 200 µl of nuclease-free water.

5.3.6.4 Plasmid Digestion

Plasmid digestion was performed to enable moving an insert from one vector to another. HA-tagged VHL was transferred from pcDNA3.1 to pIRES2-AcGFP1. A digestion mix was prepared as per Table 5.12 and then incubated for 1-3 hours at 37 ^{o}C.

Table 5.12 Digestion mix for VHL insert

Reagent	**Volume (µl)**
Multicore Buffer (x10, Promega)	3
BSA (Diluted 1/10, Promega)	3
SacI (Promega)	1
PstI (Promega)	2
Water	16
Plasmid DNA	5
TOTAL	30

5.3.6.5 Gel extraction

Gel extraction was performed to establish that the insert had been removed from the vector and to purify the insert from the vector and digestion mix before ligation. Agarose gel electrophoresis was performed by loading the total ligation mix and running at 180V for 1 hour. The insert was cut out of the gel under ultraviolet light and the DNA extracted using

the QIAquick Spin Kit (Qiagen), as per the manufacturer's instructions. Briefly, the piece of gel was weighed and placed in a 1.5ml tube. Three volumes of QG buffer were added per one volume of gel (100 µl QG per 100 mg gel). The tube was incubated at 50 ^{o}C for 10 minutes until the gel was completely dissolved. One gel volume of isopropanol was added and then the sample was added to a collection column and centrifuged at 15,600 x g for 1 minute at room temperature and the flow though discarded. Next, 0.5 ml of OG buffer was added to the column to remove any remaining agarose and the column centrifuged again and the flow through discarded. The DNA was washed by addition of 0.75 ml of PE buffer and repeat centrifugation; the flow through discarded. Finally the DNA was eluted by addition of 30 µl of nuclease free water to the centre of the column, incubation for 1 minute and centrifugation at 15600 x g for 1 minute, followed by collection of the flow though containing the plasmid insert DNA.

5.3.6.6 Ligation

Ligation of the insert into a new vector was achieved by digesting the new vector in the same way as the insert and ligating in the insert. A ligation mix was prepared as per Table 5.13. This was then incubated at 4 ^{o}C for 16 hours. The ligated vector and insert were then ready for transformation into competent cells.

Table 5.13 Ligation mix for introducing an insert into a new vector

Reagent	Volume (µl)
Buffer (x10, Promega))	2
T4 ligase (Promega)	1
Water	5
Digested vector DNA	2
Digested insert DNA	10
TOTAL	25

5.3.6.7 Site-directed mutagenesis (SDM)

SDM was achieved using the QuikChange II Site-Directed Mutagenesis Kit (Agilent). The method was performed as per the manufacturer's instructions. Briefly, PCR was carried out using a proofreading PCR enzyme with a specific thermal profile designed for the size of the plasmid being mutated (see Table 5.14.and Table 5.15). Once the PCR was completed, 0.5 µl of *DpnI* was added to each reaction, mixed and incubated at 37 ^{o}C for 1 hour, to allow digestion of the parental methylated DNA.

Table 5.14 Site directed mutagenesis PCR master mix reagents and volumes

Component	**Volume per reaction (µl)**
Reaction buffer (10 x; Agilent)	2.5
dNTP mix	0.5
Forward primer (10µM; Sigma)	0.6
Reverse primer (10µM; Sigma)	0.6
PfuUltra HF DNA Polymerase (Agilent)	0.5
dsDNA template (25ng)	1.7
ddH_2O	18.6
TOTAL	**25**

Table 5.15 Thermal cycling conditions for site directed mutagenesis of VHL in pcDNA3.1

Step	**Temperature**	**Time**	**Cycles**
1 (Initial denaturation	95 ^{o}C	30 seconds	1
2 (Denaturation)	95 ^{o}C	30 seconds	30
3 (Primer annealing)	55 ^{o}C	1 minute	
4 (Extension)	68 ^{o}C	6 minutes	
5 (Final extension)	68 ^{o}C	5 minutes	1

5.3.6.8 Primer design for SDM

For each mutation, primers were designed where both primers contained the desired mutation at the centre of the primer flanked by 10-15 bases on each side with a minimum 40% GC-content. Each primer annealed to the same sequence on opposite strands of the

plasmid. Each primer was between 25 and 45 bases in length with a melting temperature greater than 78 ^{o}C. For mismatch mutations the Tm was calculated as below:

Tm = 81.5 + 0.41(%GC) – 675/N - % mismatch

Where N is the primer length (bases).

5.3.6.9 Transformation following SDM

0.5 µl of a *DpnI* digested product of SDM was added to 25 µl of thawed ultra-competent cells and incubated on ice for 30 minutes. Tubes were heat shocked at 42 ^{o}C for 30 seconds, then incubated on ice for 2 minutes. 250 µl of pre-warmed SOC medium (Invitrogen) was added to each tube, and the tubes were incubated at 37 ^{o}C for one hour whilst rotating at 220 rpm. For each tube, 100 µl of cells and 150 µl of cells respectively were spread onto two separate kanamycin plates which were incubated at 37 ^{o}C for 16 hours.

5.3.7 Tissue culture techniques

All cell culture was performed in a Class II externally vented hood using sterile equipment and reagents. All reagents were stored at 4 ^{o}C and warmed to 37 ^{o}C in a water bath prior to use, unless otherwise stated. All work was performed using the human embryonic kidney cell line 293 (HEK293).

5.3.7.1 Maintenance of growing cell lines

Cells were grown in 75 cm^2 flasks stored in humidified incubators in a 95% air and 5% carbon dioxide environment at 37 ^{o}C. The growing cells were maintained in Dulbecco's modified eagle medium (DMEM; Sigma) supplemented with 10% foetal bovine serum (FBS), 200mM L-glutamine (GIBCO), penicillin/streptomycin (penicillin 10.000units/ml,

streptomycin 10.000µg/ml; GIBCO) and non-essential amino acids (GIBCO). Passage of adherent cells was achieved by removing all the medium and washing the cells with PBS. Then 1 ml of trypsin (Sigma) was added and the flask incubated at 37 ^{o}C for 5 minutes, also called trypsinisation. The cells were then checked under a light microscope to confirm they had detached. 9 ml of medium was added to give a final volume of 10 ml in the flask. A 1/10 dilution was performed by removing 9mls of medium containing suspended cells and adding 9 ml of fresh medium. Finally the flask was returned to the incubator.

5.3.7.2 Cryopreservation of cells

Cells were trypsinised (as described in 5.3.7.1), 9 ml of medium added to resuspend the cells and then the medium removed to a 15 ml tube and centrifuged at 1250 rpm for 3 minutes. The medium was removed and 1ml of freezing medium (900µl FBS; 100µl Dimethyl sulfoxide (DMSO; Sigma)) added to the cell pellet. The cells were resuspended in the freezing medium and transferred to a cryovial. The cryovial was placed in a freezing container (Nalgene), which was in turn put in a -80 ^{o}C freezer. This allows a -1 °C/minute cooling rate. Once frozen the cryovial was transferred to a liquid nitrogen storage tank.

5.3.7.3 Resurrecting cells from liquid nitrogen

Cells in a cryovial were thawed at 37 ^{o}C. The thawed cells were added to a 75 cm^2 flask and 10 ml of supplemented DMEM was added dropwise. The flask was then placed in the incubator to grow. The medium was changed, as required, to promote growth.

5.3.7.4 Assessing cell concentration

A haemocytometer was used to estimate the number of cells in a given quantity of medium. It was first cleaned with 70% ethanol and a cover slip was placed on top. 25 µl of trypsinised

cells in DMEM were added beneath the cover slip. The number of cells in four 1 mm^2 grid sections was counted using a light microscope and then the average number calculated. Given that 1 mm^2 of the grid contains 100 nl of cells, the number of cells in 1 ml was calculated by multiplying the average of the counts by 10000 (10^4).

5.3.7.5 Transient transfection

Prior to transfection, cells were transferred into 6-well tissue culture plates. First the cells were trypsinised and counted. Then, $4x10^5$ cells were added to each well with 2ml of medium. The cells were grown in the incubator until they reached 90% confluency, changing the medium if required. Transfection was performed by preparing a transfection solution containing 100 µl of Optimem (Invitrogen), 2 µg of the plasmid DNA of interest and 7 µl of Fugene (Roche), which was vortexed and incubated for 15 minutes at room temperature. Individual mixes were created for each plasmid transfection to be performed. Each solution was then added dropwise to a well in the 6-well plate. The cells were then incubated until confluent and the cells lysed to extract the protein (see 5.3.7.6).

5.3.7.6 Preparation of cell lysates from adherent cells

Tissue culture medium was removed and the adherent cells washed with 2 ml of cold PBS. Cells were lysed using 50 µl of RIPA buffer (25mM Tris-HCl pH 7.6, 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS (Thermo scientific)). Following incubation on ice for 5 minutes, cells were detached from the well using a sterile plastic scraper and removed to a 1.5 ml plastic tube. Harvested cells were incubated on ice for 10 minutes and then centrifuged at 13, 000 rpm for 20 minutes at 4 ^oC. The supernatant was transferred to a fresh tube and stored at -80 ^oC for downstream applications.

5.3.8 Protein Analysis

5.3.8.1 Protein quantification

The amount of protein obtained from each cell lysate was quantified using the DC protein assay (Biorad). 5 µl of each sample was added to separate wells of a microtitre plate, in duplicate. 25 µl of reagent A' was added to each well, followed by 200 µl of reagent B. The plate was gently agitated to mix the samples, and then left for 15 minutes. The 750 nm absorbencies were read using a Victor X3 Multilabel plate reader (PerkinElmer) and the protein concentration for each lysate was determined by comparison with a standard calibration curve prepared from known quantities of bovine serum albumin (0-2 mg/ml).

5.3.8.2 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Gel casting plates were cleaned with 70% ethanol before use and assembled in a stand. A mix for two 12% (w/v) resolving SDS-PAGE gels was prepared as per Table 5.16. 250 µl of 10% ammonium persulphate (APS, Sigma) and 5 µl tetramethyl ethylene diamine (TEMED, Sigma) was added to instigate polymerisation and the gel was immediately poured between the casting plates in the stand to 4/5 capacity. Water was added to fill the plates to ensure a straight edge. Once the gel polymerised the water was removed and the stacking gel prepared as per Table 5.17 and then 180 µl of APS and 10 µl of TEMED added. The gel was then poured to fill the casting plates, a comb inserted and the gel left to polymerise.

Table 5.16 Mix for 12% resolving gel

Reagent	Volume (ml)
Water	6.6
1.5 M TRIS-HCl pH 8.8 (Sigma)	5.0
30% Acrylamide/Bisacrylamide (37.5:1)	8.0
10% SDS	0.2

Table 5.17 Mix for stacking gel

Reagent	Volume (ml)
Water	6.8
0.5 M TRIS-HCl pH 6.8 (Sigma)	3.0
30% Acrylamide/Bisacrylamide (37.5:1)	2.0
10% SDS	0.12

Once the gels were set, they were placed into an electrophoresis tank part filled with 1x SDS running buffer (25mM Tris, 192mM Glycine, 0.1% (w/v) SDS pH8.3).

10ul of samples that had been normalised to 1.5 µg/µl were diluted in an equal volume of Laemmli sample buffer (25% v/v glycerol, 62.5mM Tris pH 6.8, 5% β-mercaptoethanol, 2% w/v SDS and 0.01% w/v bromophenol blue), boiled for 5 minutes and loaded onto the gel. Pre-stained molecular weight markers were also loaded. The samples were electrophoresed until the bromophenol blue reached the end of the gel, which typically took 2 hours at 100V.

5.3.8.3 Immunoblotting

Four pieces of filter paper and two sponges were soaked in 1 x transfer buffer (25mM Tris, 192mM glycine and 20% methanol). A sheet of Hybond membrane (GE Healthcare) was cut to size and activated by soaking in methanol. The gel was cut to remove the stacking section and the stack for electroblotting was assembled as in Figure 5.5.

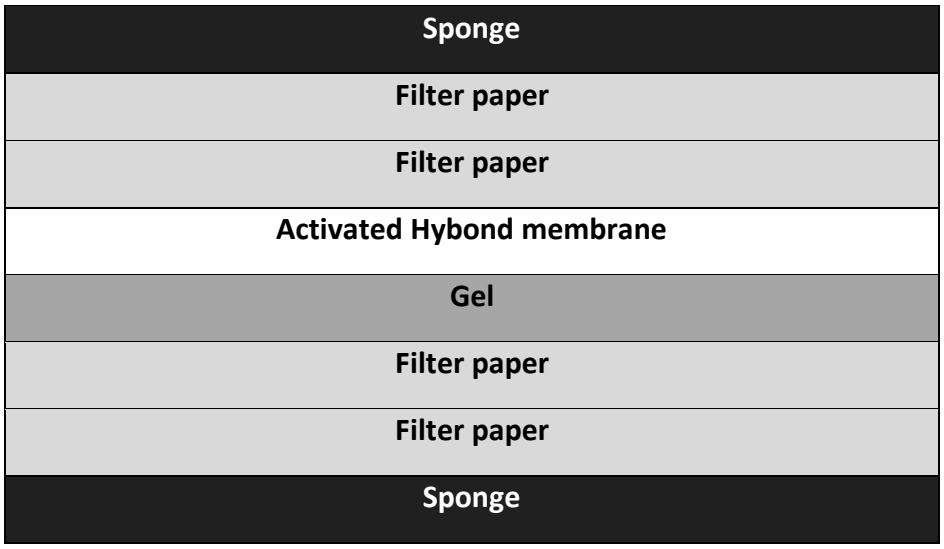


Figure 5.5 Stack for electroblotting

The stack was rolled to get rid of any air bubbles which disrupt efficient transfer. The cassette was then placed in a transfer tank containing an ice block. The cassette was orientated with the gel closest to the anode and run at 100 V for 2 hours.

Once transfer was complete, the membrane was blocked in 15 ml of milk (5% w/v dried milk (Marvel) in distilled water) for 1 hour at RT whist being agitated. The membrane was then transferred to a 25 ml polypropanyl tube containing the primary antibody (see 5.3.8.4) and a milk solution (5% w/v dried milk, 0.6% w/v bovine serum albumin (BSA) in distilled water). It was then incubated for 16 hours at 4 ^{o}C whilst on a roller. The following morning, the membrane was washed three times for five minutes in PBS-tween. The membrane was incubated in the primary antibody (see 5.3.8.4) and a milk solution (5% w/v dried milk, 0.6% w/v bovine serum albumin (BSA) in distilled water) on a roller for 1 hour. Following three more PBS-tween washes, the blot was ready to be developed.

5.3.8.4 Visualisation of proteins

The proteins were visualised using the Amersham ECL Western Blotting Detection Reagent according to the manufacturer's instructions. The membrane was then covered in plastic wrap and taped into a film cassette. In the dark room, the membrane was exposed to several photographic films, BioMax XAR (Kodak) which was developed at regular time intervals until a good quality signal was produced.

5.3.8.5 Stripping the antibodies to allow re-probing

The antibodies were stripped from a blot by boiling the membrane in distilled water for 5 minutes.

5.3.8.6 Antibodies

Table 5.18 provides details of the optimal antibody dilution for immunoblotting to obtain clear images.

Table 5.18 Dilution factors for antibodies used

Antibody	**Dilution in milk**
Primary	
HA mouse	1 in 10000
α-tubulin mouse	1 in 10000
GFP	1 in 10000
Secondary	
Rabbit anti mouse	1 in 10000

5.4 Results

5.4.1 Variants observed during gene screening analyses

A total of 75 variants (excluding Class I polymorphisms) were identified during gene screening for this research, as detailed in Chapters 3 and 4. The 75 variants included 72 heterozygous variants (see Table 5.19) and 3 mosaic variants (see Table 5.20). In the context of this study, sequence changes were considered to be a Class I polymorphisms if they: 1) had been identified in greater than 1% of a super population on either of the large variation databases available at the time of analysis (1000 genomes and Exome Sequencing Project); and/or 2) been identified *in trans* with a pathogenic mutation. In addition, variants which had been observed in one or more control populations at greater than 1% and were recorded in the literature as being putatively disease causing, were also included (n=2).

A variant was considered pathogenic in a tumour suppressor gene if it: 1) was a nonsense mutation, a frameshift mutation, at a splice site (±2 bp of CDS) or large deletion mutation (i.e. caused clear loss of function); 2) a missense, synonymous, intronic, 5'UTR or 3'UTR mutation with extensive published or local data to confirm pathogenicity. However, there were exceptions to these rules, in particular with regard to *SDHA* (see section 5.5 Discussion).

Of the 72 variants described in Table 5.19, 20 had been classified as either pathogenic or likely pathogenic and therefore clinically actionable. The remaining 52 had been classified as either likely benign or variants of unknown significance (VUS). These can be broken down by gene and their location within that gene, see Table 5.21.

Table 5.19 Heterozygous sequence variants identified during the analyses for this thesis

Complete list of the 72 sequence variants identified during gene screening for this thesis, not included those designated to be a Class I polymorphism. RCC = Renal Cell Carcinoma; PPH = Phaeochromocytoma / Paraganglioma / Head and neck paraganglioma; VUS = Variant of unknown significance.

No.	Panel	Gene	Gene location (Exon/Intron)	Variant type	Variant (codingDNA)	Variant (protein)	Times observed	Variant description (Variant class)
1	RCC	*FLCN*	7	Missense	c.[715C>T];[=]	p.(Arg239Cys)	2	VUS (III)
2	RCC	*FLCN*	11	Frameshift	c.[1285dupC];[=]	p.(His429Profs*27)	1	Pathogenic (V)
3	RCC	*FLCN*	12	Frameshift	c.[1333_1337dupGCACG];[=]	p.(Ser447Hisfs*23)	1	Pathogenic (V)
4	RCC	*FLCN*	12	Missense	c.[1333G>A];[=]	p.(Ala445Thr)	1	VUS (III)
5	RCC	*FLCN*	3'UTR	3'UTR	c.[*4A>G];[=]		1	VUS (III)
6	PPH	*MAX*	4	Nonsense	c.[223C>T];[=]	p.(Arg75*)	1	Pathogenic (V)
7	PPH	*MAX*	5	Missense	c.[425C>T];[=]	p.(Ser142Leu)	1	VUS (III)
8	RCC	*MET*	17	Missense	c.[3356G>C];[=]	p.(Gly1119Ala)	1	VUS (III)
9	PPH	*SDHA*	2	Missense	c.[136A>G];[=]	p.(Lys46Glu)	1	VUS (III)
10	PPH	*SDHA*	2	Nonsense	c.[91C>T];[=]	p.(Arg31*)	2	Likely pathogenic (IV)
11	PPH	*SDHA*	5	Synonymous	c.[549C>T];[=]	p.(=)	1	VUS (III)
12	PPH	*SDHA*	7	Synonymous	c.[822C>T];[=]	p.(=)	1	VUS (III)
13	PPH	*SDHA*	8	Synonymous	c.[1002G>A];[=]	p.(=)	1	VUS (III)
14	PPH	*SDHA*	8	Missense	c.[923C>T];[=]	p.(Thr308Met)	1	VUS (III)
15	PPH	*SDHA*	10	Missense	c.[1273G>A];[=]	p.(Val425Met)	1	VUS (III)
16	PPH	*SDHA*	10	Frameshift	c.[1338delA];[=]	p.(His447Metfs*23)	1	VUS (III)
17	PPH	*SDHA*	12	Synonymous	c.[1623G>A];[=]	p.(=)	1	VUS (III)
18	PPH	*SDHA*	13	Missense	c.[1753C>T];[=]	p.(Arg585Trp)	1	Likely pathogenic (IV)
19	PPH	*SDHA*	13	Synonymous	c.[1776T>C];[=]	p.(=)	1	VUS (III)
20	PPH	*SDHAF2*	3	Missense	c.[319C>T];[=]	p.(Arg107Cys)	1	VUS (III)

No.	Panel	Gene	Gene location (Exon/Intron)	Variant type	Variant (codingDNA)	Variant (protein)	Times observed	Variant description (Variant class)
21	PPH	*SDHB*	1	Deletion	c.[-?_200+?del];[=]		1	Pathogenic (V)
22	RCC	*SDHB*	1	Missense	c.[32G>A];[=]	p.(Arg11His)	1	VUS (III)
23	PPH	*SDHB*	1	Splicing	c.[72+1G>T];[=]		1	Pathogenic (V)
24	PPH	*SDHB*	2	Missense	c.[118A>G];[=]	p.(Lys40Glu)	1	Pathogenic (V)
25	PPH	*SDHB*	2	Deletion	c.[73-?_200+?del];[=]		1	Pathogenic (V)
26	PPH	*SDHB*	2 to 7	Deletion	c.[73-?_765+?del];[=]		1	Pathogenic (V)
27	PPH	*SDHB*	4	Splicing	c.[423+1G>A];[=]		1	Pathogenic (V)
28	PPH & RCC	*SDHB*	5	Missense	c.[487T>C];[=]	p.(Ser163Pro)	6	VUS (III)
29	PPH	*SDHB*	6	Missense	c.[587G>A];[=]	p.(Cys196Tyr)	1	Pathogenic (V)
30	RCC	*SDHB*	7	Missense	c.[724C>T];[=]	p.(Arg242Cys)	1	Pathogenic (V)
31	PPH	*SDHB*	7	Missense	c.[725G>A];[=]	p.(Arg242His)	1	Pathogenic (V)
32	PPH	*SDHB*	8	Frameshift	c.[770dupT];[=]	p.(Asn258Glufs*17)	1	Likely pathogenic (IV)
33	PPH	*SDHC*	5'UTR	5'UTR	c.[-118_-117delAG];[=]		1	VUS (III)
34	PPH	*SDHC*	2	Splicing	c.[77+2dupT];[=]		1	Pathogenic (V)
35	PPH	*SDHC*	3	Missense	c.[148C>T];[=]	p.(Arg50Cys)	1	VUS (III)
36	PPH	*SDHC*	4	Missense	c.[214C>T];[=]	p.(Arg72Cys)	1	VUS (III)
37	PPH	*SDHC*	5	Missense	c.[380A>G];[=]	p.(His127Arg)	1	VUS (III)
38	PPH	*SDHD*	1	Missense	c.[34G>A];[=]	p.(Gly12Ser)	3	VUS (III)
39	PPH	*SDHD*	3	Missense	c.[242C>T];[=]	p.(Pro81Leu)	2	Pathogenic (V)
40	PPH	*SDHD*	3	Synonymous	c.[312C>T];[=]	p.(=)	1	VUS (III)
41	PPH	*TMEM127*	3	Missense	c.[268G>A];[=]	p.(Val90Met)	1	VUS (III)
42	PPH	*TMEM127*	3	Frameshift	c.[512delTinsGCC];[=]	p.(Val171Glyfs*137)	1	Pathogenic (V)
43	PPH	*TMEM127*	4	Synonymous	c.[411T>A];[=]	p.(=)	1	VUS (III)
44	PPH	*TMEM127*	4	Synonymous	c.[534C>T];[=]	p.(=)	1	VUS (III)

No.	Panel	Gene	Gene location (Exon/Intron)	Variant type	Variant (codingDNA)	Variant (protein)	Times observed	Variant description (Variant class)
45	VHL	*VHL*	5'UTR	5'UTR	c.[-3933C>T];[=]		2	VUS (III)
46	VHL	*VHL*	5'UTR	5'UTR	c.[-3197_-3195delCTC];[=]		1	VUS (III)
47	VHL	*VHL*	5'UTR	5'UTR	c.[-963G>A];[=]		1	VUS (III)
48	VHL	*VHL*	1	Synonymous	c.[183C>G];[=]	p.(Pro61Pro)	1	VUS (III)
49	RCC	*VHL*	1	Missense	c.[233A>G];[=]	p.(Asn78Ser)	1	Pathogenic (V)
50	VHL	*VHL*	1	Missense	c.[242C>T];[=]	p.(Pro81Leu)	1	VUS (III)
51	VHL	*VHL*	1	Missense	c.[338G>A];[=]	p.(Arg113Gln)	1	VUS (III)
52	VHL	*VHL*	1	Intronic	c.[340+203G>A];[=]		1	VUS (III)
53	VHL	*VHL*	1	Intronic	c.[340+280T>G];[=]		1	Likely benign (II)
54	VHL	*VHL*	1	Intronic	c.[340+376C>A];[=]		1	Likely benign (II)
55	VHL	*VHL*	1	Intronic	c.[341-123G>T];[=]		1	VUS (III)
56	VHL	*VHL*	1	Intronic	c.[341-21_341-17delAACCT];[=]		1	VUS (III)
57	VHL	*VHL*	2	Missense	c.[361G>C];[=]	p.(Asp121His)	1	VUS (III)
58	VHL	*VHL*	2	Intronic	c.[463+8C>T];[=]		1	VUS (III)
59	VHL	*VHL*	2	Intronic	c.[464-10G>A];[=]		1	VUS (III)
60	VHL	*VHL*	2	Intronic	c.[464-1212A>G];[=]		1	Likely benign (II)
61	VHL	*VHL*	2	Intronic	c.[464-1434C>T];[=]		2	Likely benign (II)
62	VHL	*VHL*	2	Intronic	c.[464-1530C>T];[=]		1	VUS (III)
63	RCC	*VHL*	3	Splice-site	c.[464-2A>G];[=]		1	Pathogenic (V)
64	VHL	*VHL*	3	Missense	c.[554A>G];[=]	p.(Tyr185Cys)	1	VUS (III)
65	RCC	*VHL*	3	Missense	c.[629G>A];[=]	p.(Arg210Gln)	1	VUS (III)
66	VHL	*VHL*	3	Synonymous	c.[639T>C];[=]	p.(Asp213Asp)	1	VUS (III)
67	VHL	*VHL*	3'UTR	3'UTR	c.[*820A>G];[=]		1	Likely benign (II)
68	VHL	*VHL*	3'UTR	3'UTR	c.[*2511A>G];[=]		1	VUS (III)

No.	Panel	Gene	Gene location (Exon/Intron)	Variant type	Variant (codingDNA)	Variant (protein)	Times observed	Variant description (Variant class)
69	VHL	*VHL*	3'UTR	3'UTR	c.[*3021T>C];[=]		1	VUS (III)
70	VHL	*VHL*	3'UTR	3'UTR	c.[*3082C>T];[=]		1	VUS (III)
71	VHL	*VHL*	3'UTR	3'UTR	c.[*3170G>A];[=]		1	VUS (III)
72	VHL	*VHL*	3'UTR	3'UTR	c.[*3482dupA];[=]		1	Likely benign (II)

Table 5.20 Mosaic sequence variants identified during the analyses for this thesis

No.	Panel	*Gene*	Gene location (Exon/Intron)	Variant type	Variant (codingDNA)	Variant (protein)	Times observed	Variant description (Variant class)
1	VHL	*VHL*	1	Missense	c.[277G=/277G>C];[=]	p.(Gly93Arg)	1	Likely pathogenic (IV)
2	VHL	*VHL*	2	Missense	c.[394C=/394C>T];[=]	p.(Gln132Ter)	1	Pathogenic (V)
3	VHL	*VHL*	3	Missense	c.[499C=/499C>T];[=]	p.(Arg167Trp)	1	Pathogenic (V)

Table 5.21 Summary of genes and types of Class II and III variants identified during full gene screening

		Variant type						
		Frameshift	**Missense**	**Synonymous**	**Intronic**	**5'UTR**	**3'UTR**	**TOTAL**
Gene	***FLCN***		2				1	**3**
	MAX		1					**1**
	MET		1					**1**
	SDHA	1	3	5				**9**
	SDHAF2		1					**1**
	SDHB		2					**2**
	SDHC		3			1		**4**
	SDHD		1	1				**2**
	TMEM127		1	2				**3**
	VHL		5	2	10	3	6	**26**
	TOTAL	**1**	**20**	**10**	**10**	**4**	**7**	**52**

5.4.2 Interrogation of variation databases

All variants designated to be Class II or III during the gene screening chapters were compared to the current versions of the 1000 genomes data in Ensembl, Exome Sequencing Project (ESP) data and Exome Aggregation Consortium (ExAC) data, as of July 2015 (see Table 5.22).

The interrogation of currently available data has allowed the reclassification of five VUSs to Class I - Benign, as a result of new data showing these variants to be present in greater than 1% of at least one super population of control individuals. The reclassified variants comprised four synonymous changes and one deep intronic variant. An additional variant, *TMEM127* c.534C>T p.(=), was found at 1.02% in the SAS population in 1000 genomes but only 0.61% in the ExAC data, which includes the 1000 genomes data. Following the criteria given here, this variant should not be downgraded to a benign polymorphism, but it would be reasonable to downgrade it to Class II. This left 47 variants for further analysis.

In addition, 2 variants, *SDHB* c.487T>C p.(Ser163Pro) and *SDHD* c.34G>A p.(Gly12Ser), which were both present in more than one population at greater than one percent, were putative disease causing mutations, and therefore warranted further investigation (246;247). These variants are discussed in depth in section 5.6 .

Table 5.22 Updated variation database minor allele frequencies (MAF) as percentages for identified Class II and III variants

Variants highlighted in purple are those which can be reclassified as a Class I polymorphism as the result of the current MAFs (criteria: found at greater than one percent in at least one super population). The three databases examined are: 1) 1000 genomes (super populations: African (AFR); Ad-Mixed American (AMR); East Asian (EAS); European (EUR) and South Asian (SAS)); 2) Exome sequencing project (ESP) (populations: European Americans (EA) and African Americans (AA)); 3) Exome Aggregation Consortium (ExAC) (populations: African (AFR); Ad-Mixed American (AMR); East Asian (EAS); South Asian (SAS); Non-Finnish European (NFE); Finnish European (FIN) and Other (OTH)).

Gene	Variant (codingDNA)	Variant (protein)	Original Variant Class	1000 genomes (MAF-%)						ESP (MAF-%)		ExAC (MAF - %)							
				G	AFR	AMR	EAS	EUR	SAS	AA	EA	**ALL**	AFR	AMR	EAS	SAS	NFE	FIN	OTH
FLCN	c.715C>T	p.(Arg239Cys)	VUS (III)	0.06	0.00	0.14	0.00	0.20	0.00	0.02	0.06	**0.02**	0.02	0.01	0.00	0.01	0.03	0.00	0.22
FLCN	c.1333G>A	p.(Ala445Thr)	VUS (III)	0.08	0.00	0.14	0.00	0.30	0.00	-	-	**0.27**	0.05	0.13	0.00	0.33	0.37	0.08	0.22
FLCN	c.*4G>A		VUS (III)	-	-	-	-	-	-	-	-	**0.00**	0.01	0.00	0.00	0.01	0.00	0.00	0.00
MAX	c.425C>T	p.(Ser142Leu)	VUS (III)	-	-	-	-	-	-	-	-	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MET	c.3356G>C	p.(Gly1119Ala)	VUS (III)	0.02	0.00	0.14	0.00	0.00	0.00	-	-	**0.01**	0.00	0.01	0.00	0.00	0.01	0.00	0.00
SDHA	c.136A>G	p.(Lys46Glu)	VUS (III)	0.06	0.23	0.00	0.00	0.00	0.00	0.41	0.00	**0.04**	0.43	0.01	0.00	0.01	0.00	0.00	0.00
SDHA	c.549C>T	p.(=)	VUS (III)	0.92	2.87	1.01	0.00	0.10	0.00	2.54	0.07	**0.27**	2.45	0.30	0.00	0.00	0.06	0.00	0.22
SDHA	c.822C>T	p.(=)	VUS (III)	1.00	2.95	1.15	0.00	0.30	0.00	2.52	0.36	**0.46**	2.53	0.38	0.00	0.06	0.36	0.00	0.22
SDHA	c.1002G>A	p.(=)	VUS (III)	0.04	0.00	0.00	0.00	0.20	0.00	0.02	0.12	**0.08**	0.02	0.05	0.00	0.01	0.13	0.00	0.00
SDHA	c.923C>T	p.(Thr308Met)	VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SDHA	c.1273G>A	p.(Val425Met)	VUS (III)	0.04	0.00	0.29	0.00	0.00	0.00	-	-	**0.00**	0.00	0.02	0.00	0.01	0.00	0.00	0.00
SDHA	c.1338delA	p.(His447Metfs*23)	VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SDHA	c.1623G>A	p.(=)	VUS (III)	0.20	0.00	0.14	0.00	0.89	0.00	0.07	0.28	**0.21**	0.08	0.19	0.00	0.10	0.29	0.12	0.44
SDHA	c.1776T>C	p.(=)	VUS (III)	1.18	4.01	0.72	0.00	0.10	0.00	4.40	0.08	**0.41**	4.45	0.12	0.00	0.02	0.02	0.00	0.22
SDHAF2	c.319C>T	p.(Arg107Cys)	VUS (III)	0.06	0.23	0.00	0.00	0.00	0.00	0.34	0.00	**0.02**	0.28	0.00	0.00	0.00	0.00	0.00	0.11
SDHB	c.32G>A	p.(Arg11His)	VUS (III)	0.14	0.53	0.00	0.00	0.00	0.00	0.68	0.00	**0.05**	0.55	0.03	0.00	0.00	0.00	0.00	0.13
SDHB	c.487T>C	p.(Ser163Pro)	VUS (III)	0.94	0.15	0.58	0.00	1.69	2.45	-	-	**1.25**	0.16	0.34	0.00	1.76	1.52	2.33	0.77
SDHC	c.-118_-117delAG		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Gene	Variant (codingDNA)	Variant (protein)	Original Variant Class	1000 genomes (MAF-%)						ESP (MAF-%)		ExAC (MAF - %)							
				G	AFR	AMR	EAS	EUR	SAS	AA	EA	**ALL**	AFR	AMR	EAS	SAS	NFE	FIN	OTH
SDHC	c.148C>T	p.(Arg50Cys)	VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SDHC	c.214C>T	p.(Arg72Cys)	VUS (III)	-	-	-	-	-	-	-	-	**0.00**	0.00	0.00	0.01	0.00	0.00	0.00	0.00
SDHC	c.380A>G	p.(His127Arg)	VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SDHD	c.34G>A	p.(Gly12Ser)	VUS (III)	0.50	0.08	1.59	0.10	0.99	0.20	-	-	**0.73**	0.21	0.81	0.01	0.27	1.02	0.48	1.11
SDHD	c.312C>T	p.(=)	VUS (III)	0.70	2.57	0.14	0.00	0.00	0.00	1.93	0.01	**0.21**	2.39	0.10	0.00	0.00	0.00	0.00	0.11
TMEM127	c.268G>A	p.(Val90Met)	VUS (III)	0.18	0.68	0.00	0.00	0.00	0.00	0.82	0.00	**0.08**	0.84	0.05	0.00	0.00	0.00	0.00	0.11
TMEM127	c.411T>A	p.(=)	VUS (III)	-	-	-	-	-	-	-	-	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.11
TMEM127	c.534C>T	p.(=)	VUS (III)	0.20	0.00	0.00	0.00	0.00	1.02	-	-	**0.08**	0.00	0.01	0.00	0.61	0.00	0.00	0.00
VHL	c.-3933C>T		VUS (III)	0.04	0.00	0.00	0.00	0.10	0.10	-	-	-	-	-	-	-	-	-	-
VHL	c.-3197_-3195delCTC		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.-963G>A		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.183C>G	p.(Pro61Pro)	VUS (III)	0.04	0.00	0.00	0.00	0.00	0.20	0.00	0.08	**0.25**	0.00	0.34	0.15	0.47	0.20	0.00	0.00
VHL	c.242C>T	p.(Pro81Leu)	VUS (III)	-	-	-	-	-	-	-	-	**0.00**	0.00	0.00	0.00	0.01	0.00	0.00	0.00
VHL	c.338G>A	p.(Arg113Gln)	VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.340+203G>A		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.340+280T>G		Likely benign (II)	0.04	0.00	0.00	0.00	0.20	0.00	-	-	-	-	-	-	-	-	-	-
VHL	c.340+376C>A		Likely benign (II)	0.02	0.00	0.00	0.00	0.10	0.00	-	-	-	-	-	-	-	-	-	-
VHL	c.341-123G>T		VUS (III)	0.14	0.00	0.00	0.00	0.00	0.72	-	-	-	-	-	-	-	-	-	-
VHL	c.341-21_341-17delAACCT		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.361G>C	p.(Asp121His)	VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.463+8C>T		VUS (III)	0.02	0.00	0.00	0.00	0.10	0.00	0.11	0.07	**0.05**	0.05	0.00	0.00	0.00	0.08	0.00	0.00
VHL	c.464-10G>A		VUS (III)	-	-	-	-	-	-	-	-	**0.00**	0.01	0.00	0.00	0.00	0.00	0.00	0.00
VHL	c.464-1212A>G		Likely benign (II)	0.44	0.00	0.72	0.00	0.60	1.12	-	-	-	-	-	-	-	-	-	-

Gene	Variant (codingDNA)	Variant (protein)	Original Variant Class	1000 genomes (MAF-%)						ESP (MAF-%)		ExAC (MAF - %)							
				G	AFR	AMR	EAS	EUR	SAS	AA	EA	**ALL**	AFR	AMR	EAS	SAS	NFE	FIN	OTH
VHL	c.464-1434C>T		Likely benign (II)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.464-1530C>T		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.554A>G	p.(Tyr185Cys)	VUS (III)	0.02	0.00	0.00	0.00	0.10	0.00	-	-	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
VHL	c.629G>A	p.(Arg210Gln)	VUS (III)	0.02	0.08	0.00	0.00	0.00	0.00	0.00	0.01	**0.01**	0.01	0.00	0.00	0.00	0.01	0.00	0.00
VHL	c.639T>C	p.(Asp213Asp)	VUS (III)	-	-	-	-	-	-	-	-	**0.00**	0.00	0.00	0.00	0.00	0.01	0.00	0.00
VHL	c.*820A>G		Likely benign (II)	0.08	0.00	0.29	0.00	0.20	0.00	-	-	-	-	-	-	-	-	-	-
VHL	c.*2511A>G		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.*3021T>C		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.*3082C>T		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.*3170G>A		VUS (III)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VHL	c.*3482dupA		Likely benign (II)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

5.4.3 In silico analysis of variants identified

5.4.3.1 *In silico* splicing prediction

In silico prediction of splicing was performed with five different tools for the remaining 47 Class II and III variants using Alamut as an interface (see Appendix Table 7.4). Four variants (8.5%) sat within the splice consensus sequences. Two sat in the 5' sequence, one exonic and one intronic and two in the 3' sequence, one exonic and one intronic. These variants were examined using the pathway (see Figure 5.2) determined by Houdayer *et. al.* (219), which incorporates two of the five programs employed (MES and SSF). The results are given in Table 5.23. None of the four variants reached the threshold for RNA analysis.

Table 5.23 Examination of variants affecting 5'/3' consensus splice sequences using the pathway devised by Houdayer *et. al.*

Gene	***TMEM127***	***VHL***	***VHL***	***VHL***
Variant (coding DNA)	**c.411T>A**	**c.338G>A**	**c.463+8C>T**	**c.464-10G>A**
Variant (protein)	**p.(=)**	**p.(Arg113Gln)**		
Consensus sequence affected	3'	5'	5'	3'
Consensus sequence above reliability threshold?	Yes	Yes	Yes	Yes
MES at least 15% lower than WT?	No (9.8% ↓)	No (9.6% ↑)	No (No change)	No (6.8% ↓)
SSF at least 5% lower than WT?	No (No change)	No (0.9% ↑)	No (No change)	No (No change)
RNA analysis recommended?	**No**	**No**	**No**	**No**

All variants were then examined using more permissive criteria (considered to warrant further investigation if two or more programs predicted a ≥10% decrease of a true splice site or a ≥10% increase for a cryptic splice site). The results are summarised in Table 5.24. Of the 47 variants examined, one (2%) was predicted to affect an actual splice site and five (11%) were predicted to create a cryptic splice site.

Figure 5.6 shows the data for the sole variant predicted to affect a normally-used splice site, *VHL* c.341-21_341-17delAACCT. The local sequence was examined for any potential cryptic splice sites using the data from the splicing prediction algorithms; however, none were predicted. Therefore, it is hypothesised that this variant may stop or decrease the use of the VHL exon 2 splice acceptor site, leading to skipping of exon 2.

Of the five variants where two or more programs showed a ≥10% increase in scores for a cryptic splice site, four were exonic and one was within the 3'UTR of the gene. None of these variants were within the 5' or 3' consensus sequences. In order to further investigate their potential to affect splicing, the scores for these variants were compared to the average scores less two standard deviations for the actual donors or acceptors for that gene, as appropriate. A summary of these results is shown in Table 5.25. Of the five variants examined, three only equalled or exceeded the threshold (average minus 2SD) for 1 or 2 programs, but two reached or exceeded the threshold for 4 out of 5 programs. The sequence local to the variants was also examined for predicted but unused splice sites. This showed that *VHL* c.[*3082C>T] has a local predicted but unused splice site (see Figure 5.7), suggesting that RNA analysis would not be worthwhile, especially given that a mechanism of aberrant splicing so far into a 3'UTR has never been reported.

Interestingly, *TMEM127* c.534C>T p.(=) has a MAF of >1% in the SAS population in the 1000 genomes study and 0.61% in the same population in the ExAC database. So is unlikely to be pathogenic, however confirmation that it does not affect splicing would be worthwhile.

Given these finding it was concluded that RNA analysis would only be worthwhile for *VHL* c.341-21_341-17delAACCT and *TMEM127* c.534C>T p.(=).

Table 5.24 Summary of splicing predictions of 5 splicing algorithms for identified variants
An effect on splicing was considered to be indicated if the score was reduced by greater than 10% in more than one program for the actual donor or acceptor, or increased by greater than 10% in more than one program for a cryptic splice donor or acceptor.

Gene	Variant (coding DNA)	Variant (protein)	Actual donor/acceptor score reduced	Cryptic donor/acceptor score increased
FLCN	c.715C>T	p.(Arg239Cys)	No	No
FLCN	c.1333G>A	p.(Ala445Thr)	No	No
FLCN	c.*4G>A		No	No
MAX	c.425C>T	p.(Ser142Leu)	No	No
MET	c.3356G>C	p.(Gly1119Ala)	No	No
SDHA	c.136A>G	p.(Lys46Glu)	No	No
SDHA	c.923C>T	p.(Thr308Met)	No	No
SDHA	c.1002G>A	p.(=)	No	No
SDHA	c.1273G>A	p.(Val425Met)	No	No
SDHA	c.1338delA	p.(His447Metfs *23)	No	No
SDHA	c.1623G>A	p.(=)	No	No
SDHAF2	c.319C>T	p.(Arg107Cys)	No	No
SDHB	c.32G>A	p.(Arg11His)	No	No
SDHB	c.487T>C	p.(Ser163Pro)	No	No
SDHC	c.-118_-117delAG		No	No
SDHC	c.148C>T	p.(Arg50Cys)	No	No
SDHC	c.214C>T	p.(Arg72Cys)	No	No
SDHC	c.380A>G	p.(His127Arg)	No	No
SDHD	c.34G>A	p.(Gly12Ser)	No	Yes
TMEM127	c.268G>A	p.(Val90Met)	No	No
TMEM127	c.411T>A	p.(=)	No	No
TMEM127	c.534C>T	p.(=)	No	Yes
VHL	c.-3933C>T		Not available	Not available
VHL	c.-3197_-3195delCTC		Not available	Not available
VHL	c.-963G>A		No	No
VHL	c.183C>G	p.(Pro61Pro)	No	Yes
VHL	c.242C>T	p.(Pro81Leu)	No	No
VHL	c.338G>A	p.(Arg113Gln)	No	No
VHL	c.340+203G>A		No	No
VHL	c.340+280T>G		No	No
VHL	c.340+376C>A		No	No
VHL	c.341-123G>T		No	No
VHL	c.341-21_341-17delAACCT		Yes	No
VHL	c.361G>C	p.(Asp121His)	No	No
VHL	c.463+8C>T		No	No

Gene	Variant (coding DNA)	Variant (protein)	Actual donor/acceptor score reduced	Cryptic donor/acceptor score increased
VHL	c.464-10G>A		No	No
VHL	c.464-1434C>T		No	No
VHL	c.464-1530C>T		No	No
VHL	c.554A>G	p.(Tyr185Cys)	No	No
VHL	c.629G>A	p.(Arg210Gln)	No	Yes
VHL	c.639T>C	p.(Asp213Asp)	No	No
VHL	c.*820A>G		No	No
VHL	c.*2511A>G		No	No
VHL	c.*3021T>C		No	No
VHL	c.*3082C>T		No	Yes
VHL	c.*3170G>A		No	No
VHL	c.*3482dupA		No	No

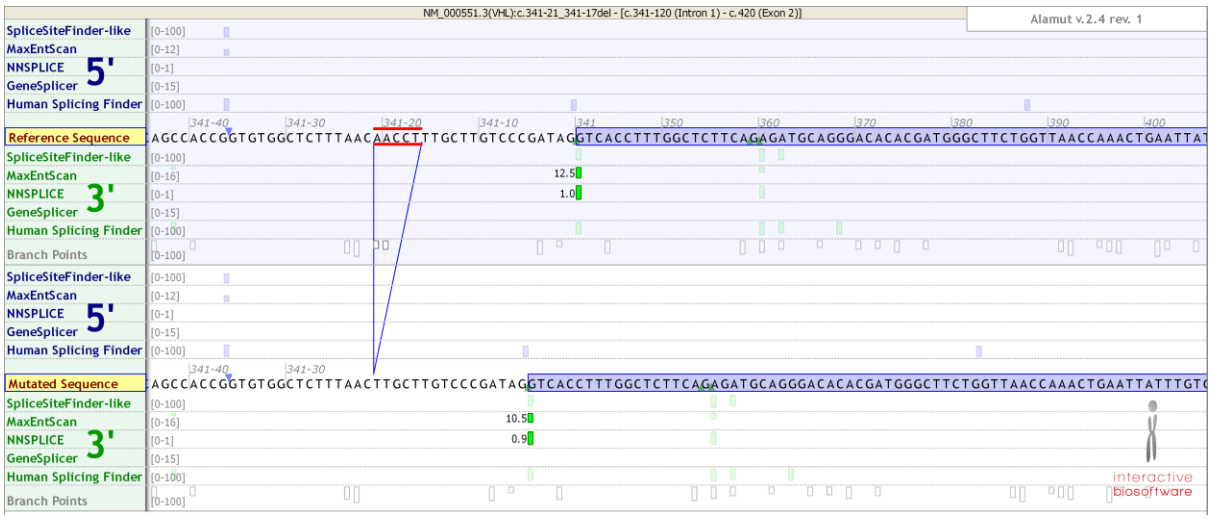


Figure 5.6 Alamut splicing prediction for the *VHL* variant c.341-21_341-17delAACCT
Four programs identified the splice acceptor. Half of the programs predicted a small drop in the score given to the sequence; the MaxEntScan score decreased by 16% and the NNSplice score by 10%. In addition, 2 predicted branch points were removed by the deletion but these were not the best predicted sites.

Table 5.25 Summary of comparison of predicted cryptic splice sites caused by variant with true splice sites for the same gene

The number of programs was recorded (maximum 5) for which the score for potential cryptic splice sites is equal to or exceeds the average scores minus two standard deviations for the comparable true splice sites, 5' or 3' considered as appropriate (full calculations available in Appendix Table 7.5).

Gene	Variant (coding DNA)	Variant (protein)	Prediction	Number of algorithms for which variant score is equal to or greater than the average scores less 2SDs of the real splice sites for that gene
SDHD	c.34G>A	p.(Gly12Ser)	3' Acceptor created	1 of 5
TMEM127	c.534C>T	p.(=)	3' Acceptor created	4 of 5
VHL	c.183C>G;	p.(Pro61Pro)	5' Donor created	2 of 5
VHL	c.629G>A	p.(Arg210Gln)	3' Acceptor created	1 of 5
VHL	c.*3082C>T		5' Donor created	4 of 5

Figure 5.7 Predicted splice sites local to *VHL* c.*3082C>T

VHL c.*3082C>T is predicted to create a splice donor site (red box), however another splice donor site (blue box) is predicted at a similar level in both the wildtype and variant sequences, which is unused in normal splicing of the gene.

5.4.3.2 In silico missense analysis

Twenty sequence variants caused missense changes in their respective proteins. One occurred in a proto-oncogene, *MET* c.3356G>C p.(Gly1119Ala), and therefore use of *in silico* analysis programs, which look for deleterious changes, was not considered worthwhile and it was excluded, leaving 19 variants for analysis.

5.4.3.2.1 In silico missense analysis - tools commonly used by diagnostic genetics laboratories

The three in silico tools most commonly used by UK diagnostic genetics laboratories were used to interrogate the variants (see Table 5.26). 7/19 variants showed concordant results for these programs, with 5 variants predicted to be benign and 2 predicted to be pathogenic. The remaining 12 variants gave conflicting results.

5.4.3.2.2 In silico missense analysis - additional in silico missense variant prediction tools

First, five sequence and evolutionary conservation-based prediction tools were run (see Table 5.27), including Align-GVGD and SIFT which were included in analyses in the section above. 11/19 variants gave concordant results for all tools for which a result was returned, with seven variants predicted to be benign and four variants predicted to be pathogenic. An additional six variants were concordant for greater than 75% of the tools for which a result was returned, with three variants predicted to be benign and three predicted to be pathogenic.

Next, eight machine learning tools were run (see Table 5.28). Only 4/19 gave 100% concordant results for all tools for which a result was returned. An additional 3/19 gave

concordant results for greater than 75% of tools which returned a result, with an additional one variant predicted to be benign and two predicted to be pathogenic.

5.4.3.2.3 In silico missense analysis – meta-analysis tools

Three meta-analysis tools were interrogated (see Table 5.29). 11/19 variants gave concordant results for these programs, with four variants predicted to be benign and seven variants predicted to be pathogenic, the remaining eight gave conflicting results.

5.4.3.2.4 In silico missense analysis – concordance between sub-groups of prediction tools

The concordance between subgroups of tools was then examined (see Table 5.30). A variant was considered to be concordant between all sub-groups if ≥75% of the tools examined for each category gave the same result. If <75% of tools were concordant for a category of tool then that category was excluded. Using this method, 5 variants were assigned as benign, 5 as pathogenic and 9 as unknown.

5.4.3.2.5 In silico missense analysis – comparison between amalgamated predictions for all tools and tools used by diagnostic laboratories.

Finally the amalgamated results for all the 17 different in silico tools were compared to the amalgamated results of the three in silico tools used by diagnostic labs (see Table 5.31). All seven predictions by the three diagnostically used tools (100% concordance) were matched by the amalgamated results (using ≥75% concordance within a program type). An additional three predictions were made by the amalgamated tools.

Table 5.26 *In silico* predictions using programs commonly in use in UK diagnostic laboratories

Red writing indicates the variant is considered to be damaging, yellow indicates the effect is unclear and green indicates the variant is predicted to be benign. Concordance between the programs is demonstrated in the last column.

Gene	Exon	Variant (coding DNA)	Variant (protein)	AlignGVGD	Polyphen-2 (HumDiv)	Polyphen-2 (HumVar)	SIFT	100% Concordant?
FLCN	7	c.715C>T	p.(Arg239Cys)	Class C65 (GV: 0.00 - GD: 179.53)	PROBABLY DAMAGING: 1.000	PROBABLY DAMAGING: 0.999	Deleterious (score: 0)	YES (Pathogenic)
FLCN	12	c.1333G>A	p.(Ala445Thr)	Class C0 (GV: 171.88 - GD: 0.00)	BENIGN: 0.000	BENIGN: 0.000	Tolerated (score: 1)	YES (Benign)
MAX	5	c.425C>T	p.(Ser142Leu)	Class C0 (GV: 181.93 - GD: 0.00)	PROBABLY DAMAGING: 0.999	PROBABLY DAMAGING: 0.984	Tolerated (score: 0.33)	NO
SDHA	2	c.136A>G	p.(Lys46Glu)	Class C0 (GV: 126.89 - GD: 43.63)	BENIGN: 0.010	BENIGN: 0.005	Tolerated (score: 0.11)	YES (Benign)
SDHA	8	c.923C>T	p.(Thr308Met)	Class C65 (GV: 0.00 - GD: 81.04)	PROBABLY DAMAGING: 1.000	PROBABLY DAMAGING: 1.000	Deleterious (score: 0)	YES (Pathogenic)
SDHA	10	c.1273G>A	p.(Val425Met)	Class C0 (GV: 251.13 - GD: 20.52)	POSSIBLY DAMAGING: 0.819	POSSIBLY DAMAGING: 0.516	Tolerated (score: 0.05)	NO
SDHAF2	3	c.319C>T	p.(Arg107Cys)	Class C15 (GV: 124.29 - GD: 109.21)	POSSIBLY DAMAGING: 0.641	BENIGN: 0.165	Deleterious (score: 0)	NO
SDHB	1	c.32G>A	p.(Arg11His)	Class C0 (GV: 241.31 - GD: 1.62)	POSSIBLY DAMAGING: 0.874	BENIGN: 0.165	Deleterious (score: 0)	NO
SDHB	5	c.487T>C	p.(Ser163Pro)	Class C0 (GV: 144.64 - GD: 0.00)	BENIGN: 0.000	BENIGN: 0.000	Tolerated (score: 0.3)	YES (Benign)
SDHC	3	c.148C>T	p.(Arg50Cys)	Class C25 (GV: 85.11 - GD: 126.72)	PROBABLY DAMAGING: 1.000	PROBABLY DAMAGING: 0.996	Deleterious (score: 0)	NO
SDHC	4	c.214C>T	p.(Arg72Cys)	Class C65 (GV: 0.00 - GD: 179.53)	PROBABLY DAMAGING: 0.995	POSSIBLY DAMAGING: 0.815	Deleterious (score: 0)	NO
SDHC	5	c.380A>G	p.(His127Arg)	Class C0 (GV: 227.83 - GD: 28.42)	PROBABLY DAMAGING: 1.000	PROBABLY DAMAGING: 1.000	Deleterious (score: 0)	NO
SDHD	1	c.34G>A	p.(Gly12Ser)	Class C0 (GV: 353.86 - GD: 0.00)	BENIGN: 0.005	BENIGN: 0.003	Tolerated (score: 0.6)	YES (Benign)
TMEM127	3	c.268G>A	p.(Val90Met)	Class C0 (GV: 28.68 - GD: 0.00)	POSSIBLY DAMAGING: 0.739	POSSIBLY DAMAGING: 0.530	Deleterious (score: 0.02)	NO
VHL	1	c.242C>T	p.(Pro81Leu)	Class C25 (GV: 37.56 - GD: 76.16)	PROBABLY DAMAGING: 0.997	POSSIBLY DAMAGING: 0.64	Deleterious (score: 0.02)	NO
VHL	1	c.338G>A	p.(Arg113Gln)	Class C0 (GV: 108.48 - GD: 16.48)	POSSIBLY DAMAGING: 0.520	BENIGN: 0.012	Tolerated (score: 0.25)	NO
VHL	2	c.361G>C	p.(Asp121His)	Class C0 (GV: 95.81 - GD: 5.07)	PROBABLY DAMAGING: 1.000	PROBABLY DAMAGING: 0.995	Deleterious (score: 0.04)	NO
VHL	3	c.554A>G	p.(Tyr185Cys)	Class C0 (GV: 257.44 - GD: 11.33)	BENIGN: 0.372	BENIGN: 0.276	Deleterious (score: 0.03)	NO
VHL	3	c.629G>A	p.(Arg210Gln)	Class C0 (GV: 238.02 - GD: 0.00)	BENIGN: 0.002	BENIGN: 0.001	Tolerated (score: 0.61)	YES (Benign)

Table 5.27 *In silico* predictions using programs categorised as sequence and evolutionary conservation-based

Red text indicates the variant is considered to be damaging, yellow indicates the effect is unclear and green indicates the variant is predicted to be benign. Concordance between the programs is demonstrated in the last column.

Gene	Exon	Variant (codingDNA)	Variant (protein)	AlignGVGD	SIFT	Mutation Assessor	PANTHER	PROVEAN	100% Concordant?	≥75% Concordant?
FLCN	7	c.715C>T	p.(Arg239Cys)	Class C65	Deleterious	medium (2.765)	Error (-)	Deleterious (-5.68)	**Yes (Pathogenic)**	-
FLCN	12	c.1333G>A	p.(Ala445Thr)	Class C0	Tolerated	neutral (-1.1)	Error (-)	Neutral (0.19)	**Yes (Benign)**	-
MAX	5	c.425C>T	p.(Ser142Leu)	Class C0	Tolerated	low (1.59)	Not significant (-1.68274)	Deleterious (-2.57)	**No**	**Yes (Benign)**
SDHA	2	c.136A>G	p.(Lys46Glu)	Class C0	Tolerated	low (1.21)	Not significant (-0.31746)	Neutral (-0.47)	**Yes (Benign)**	-
SDHA	8	c.923C>T	p.(Thr308Met)	Class C65	Deleterious	high (4.725)	Significant (-4.7476)	Deleterious (-5.61)	**Yes (Pathogenic)**	-
SDHA	10	c.1273G>A	p.(Val425Met)	Class C0	Tolerated	low (1.15)	Error (-)	Neutral (-0.33)	**Yes (Benign)**	-
SDHAF2	3	c.319C>T	p.(Arg107Cys)	Class C15	Deleterious	medium (3.025)	Significant (-4.19036)	Deleterious (-5.53)	**Yes (Pathogenic)**	-
SDHB	1	c.32G>A	p.(Arg11His)	Class C0	Deleterious	neutral (0)	Error (-)	Neutral (-0.5)	**No**	**Yes (Benign)**
SDHB	5	c.487T>C	p.(Ser163Pro)	Class C0	Tolerated	neutral (0)	Not significant (-2.04801)	Neutral (0.44)	**Yes (Benign)**	-
SDHC	3	c.148C>T	p.(Arg50Cys)	Class C25	Deleterious	high (4.11)	Significant (-6.25127)	Deleterious (-6.19)	**No**	**Yes (Pathogenic)**
SDHC	4	c.214C>T	p.(Arg72Cys)	Class C65	Deleterious	high (3.86)	Significant (-6.82403)	Deleterious (-7.07)	**Yes (Pathogenic)**	-
SDHC	5	c.380A>G	p.(His127Arg)	Class C0	Deleterious	high (4.24)	Significant (-5.71111)	Deleterious (-7.47)	**No**	**Yes (Pathogenic)**
SDHD	1	c.34G>A	p.(Gly12Ser)	Class C0	Tolerated	neutral (0.145)	Not significant (-1.27688)	Neutral (0.21)	**Yes (Benign)**	-
TMEM127	3	c.268G>A	p.(Val90Met)	Class C0	Deleterious	neutral (0.345)	Error (-)	Neutral (-1.43)	**No**	**Yes (Benign)**
VHL	1	c.242C>T	p.(Pro81Leu)	Class C25	Deleterious	low (1.445)	Significant (-3.63422)	Deleterious (-4.31)	**No**	**No**
VHL	1	c.338G>A	p.(Arg113Gln)	Class C0	Tolerated	low (0.895)	Not significant (-2.84052)	Neutral (-0.53)	**Yes (Benign)**	-
VHL	2	c.361G>C	p.(Asp121His)	Class C0	Deleterious	medium (2.435)	Significant (-3.39572)	Deleterious (-5.32)	**No**	**Yes (Pathogenic)**
VHL	3	c.554A>G	p.(Tyr185Cys)	Class C0	Deleterious	low (1.39)	Significant (-4.0583)	Deleterious (-4.59)	**No**	**No**
VHL	3	c.629G>A	p.(Arg210Gln)	Class C0	Tolerated	neutral (0)	Not significant (-1.21391)	Neutral (-0.44)	**Yes (Benign)**	-

Table 5.28 *In silico* predictions using programs categorised as machine learning-based

Red text indicates the variant is considered to be damaging and green indicates the variant is predicted to be benign. Concordance between the programs is demonstrated in the last two columns. RI = reliability index, Prob of Path = probability of pathogenicity.

Gene	Exon	Variant (codingDNA)	Variant (protein)	I-Mutant (RI)	Mutation Taster (p-value)	MutPred (Prob of Path)	nsSNP Analyzer	PON-P2 (Prob of Path)	PhD-SNP (RI)	SNAP (RI)	SNPs&GO (RI)	100% Concordant	≥75% Concordant
FLCN	7	c.715C>T	p.(Arg239Cys)	Disease (4)	Disease causing (1)	Disease (0.916)	UNKNOWN	Pathogenic (0.912)	Neutral (3)	Non-neutral (6)	Neutral (5)	No	No
FLCN	12	c.1333G>A	p.(Ala445Thr)	Disease (3)	Polymorphism (1)	Neutral (0.281)	UNKNOWN	Neutral (0.058)	Neutral (6)	Non-neutral (1)	Neutral (7)	No	No
MAX	5	c.425C>T	p.(Ser142Leu)	Neutral (8)	Disease causing (1)	Neutral (0.304)	UNKNOWN	Unknown (0.607)	Neutral (6)	Non-neutral (2)	Neutral (9)	No	No
SDHA	2	c.136A>G	p.(Lys46Glu)	Disease (3)	Disease causing (1)	Neutral (0.425)	NEUTRAL	Unknown (0.594)	Disease (0)	Neutral (1)	Neutral (2)	No	No
SDHA	8	c.923C>T	p.(Thr308Met)	Disease (4)	Disease causing (1)	Disease (0.86)	DELETERIOUS	Pathogenic (0.994)	Disease (0)	Non-neutral (5)	Disease (9)	Yes (Pathogenic)	-
SDHA	10	c.1273G>A	p.(Val425Met)	Disease (4)	Polymorphism (0.73)	Neutral (0.292)	NEUTRAL	Unknown (0.529)	Neutral (2)	Neutral (4)	Neutral (3)	No	Yes (Benign)
SDHAF2	3	c.319C>T	p.(Arg107Cys)	Disease (6)	Disease causing (1)	Neutral (0.488)	UNKNOWN	Unknown (0.843)	Disease (3)	Neutral (2)	Disease (0)	No	No
SDHB	1	c.32G>A	p.(Arg11His)	Disease (3)	Disease causing (0.99)	Disease (0.775)	UNKNOWN	Neutral (0.193)	Neutral (3)	Non-neutral (0)	Disease (3)	No	No
SDHB	5	c.487T>C	p.(Ser163Pro)	Disease (3)	Disease causing (1)	Neutral (0.14)	UNKNOWN	Unknown (0.666)	Disease (0)	Neutral (2)	Neutral (5)	No	No
SDHC	3	c.148C>T	p.(Arg50Cys)	Disease (5)	Disease causing (1)	Disease (0.971)	UNKNOWN	Pathogenic (0.878)	Disease (9)	Non-neutral (6)	Disease (9)	Yes (Pathogenic)	-
SDHC	4	c.214C>T	p.(Arg72Cys)	Disease (6)	Disease causing (1)	Disease (0.984)	UNKNOWN	Pathogenic (0.93)	Disease (9)	Non-neutral (6)	Disease (8)	Yes (Pathogenic)	-
SDHC	5	c.380A>G	p.(His127Arg)	Disease (6)	Disease causing (1)	Disease (0.91)	UNKNOWN	Pathogenic (0.952)	Disease (9)	Non-neutral (6)	Disease (9)	Yes (Pathogenic)	-
SDHD	1	c.34G>A	p.(Gly12Ser)	Disease (4)	Disease causing (0)	Neutral (0.127)	UNKNOWN	Neutral (0.036)	Neutral (8)	Neutral (2)	Neutral (6)	No	No
TMEM127	3	c.268G>A	p.(Val90Met)	Disease (2)	Disease causing (0.99)	Neutral (0.743)	UNKNOWN	Unknown (0.434)	Neutral (8)	Non-neutral (3)	Neutral (7)	No	No
VHL	1	c.242C>T	p.(Pro81Leu)	Disease (8)	Disease causing (0.98)	Neutral (0.708)	NEUTRAL	Unknown (0.744)	Disease (4)	Neutral (0)	Disease (9)	No	No
VHL	1	c.338G>A	p.(Arg113Gln)	Disease	Polymorphism (0.99)	Neutral	NEUTRAL	Pathogenic (0.838)	Neutral	Neutral (2)	Disease	No	No

Gene	Exon	Variant (codingDNA)	Variant (protein)	I-Mutant (RI)	Mutation Taster (p-value)	MutPred (Prob of Path)	nsSNP Analyzer	PON-P2 (Prob of Path)	PhD-SNP (RI)	SNAP (RI)	SNPs&GO (RI)	100% Concordant	≥75% Concordant
				(2)		(0.715)			(7)		(8)		
VHL	2	c.361G>C	p.(Asp121His)	Disease (5)	Disease causing (1)	Disease (0.842)	NEUTRAL	Pathogenic (0.867)	Disease (4)	Non-neutral (4)	Disease (9)	No	Yes (Pathogenic)
VHL	3	c.554A>G	p.(Tyr185Cys)	Disease (5)	Disease causing (0.96)	Neutral (0.689)	DELETERIOUS	Pathogenic (0.89)	Disease (2)	Non-neutral (0)	Disease (9)	No	Yes (Pathogenic)
VHL	3	c.629G>A	p.(Arg210Gln)	Neutral (4)	Polymorphism (1)	Neutral (0.525)	UNKNOWN	Unknown (0.285)	Neutral (9)	Non-neutral (0)	Disease (0)	No	No

Table 5.29 *In silico* predictions from 3 meta-predictors

Red text indicates the variant is considered to be damaging and green indicates the variant is predicted to be benign. Concordance between the programs is demonstrated in the last column.

Gene	Exon	Variant (codingDNA)	Variant (protein)	CONDEL	Meta-SNP	PredictSNP	100% Concordant?
FLCN	7	c.715C>T	p.(Arg239Cys)	D (0.63)	Disease (0.743, 5)	DELETERIOUS (0.76)	Yes (Pathogenic)
FLCN	12	c.1333G>A	p.(Ala445Thr)	N (0.4)	Neutral (0.063, 9)	NEUTRAL (0.83)	Yes (Benign)
MAX	5	c.425C>T	p.(Ser142Leu)	D (0.57)	Neutral (0.165, 7)	DELETERIOUS (0.55)	No
SDHA	2	c.136A>G	p.(Lys46Glu)	D (0.52)	Neutral (0.297, 4)	NEUTRAL (0.75)	No
SDHA	8	c.923C>T	p.(Thr308Met)	D (0.77)	Disease (0.82, 6)	DELETERIOUS (0.87)	Yes (Pathogenic)
SDHA	10	c.1273G>A	p.(Val425Met)	N (0.51)	Neutral (0.242, 5)	NEUTRAL (0.61)	Yes (Benign)
SDHAF2	3	c.319C>T	p.(Arg107Cys)	D (0.69)	Disease (0.614, 2)	NEUTRAL (0.6)	No
SDHB	1	c.32G>A	p.(Arg11His)	D (0.53)	Neutral (0.401, 2)	NEUTRAL (0.68)	No
SDHB	5	c.487T>C	p.(Ser163Pro)	N (0.5)	Neutral (0.433, 1)	NEUTRAL (0.74)	Yes (Benign)
SDHC	3	c.148C>T	p.(Arg50Cys)	D (0.73)	Disease (0.854, 7)	DELETERIOUS (0.87)	Yes (Pathogenic)
SDHC	4	c.214C>T	p.(Arg72Cys)	D (0.72)	Disease (0.823, 6)	DELETERIOUS (0.87)	Yes (Pathogenic)
SDHC	5	c.380A>G	p.(His127Arg)	D (0.79)	Disease (0.888, 8)	DELETERIOUS (0.87)	Yes (Pathogenic)
SDHD	1	c.34G>A	p.(Gly12Ser)	N (0.51)	Neutral (0.1, 8)	NEUTRAL (0.83)	Yes (Benign)
TMEM127	3	c.268G>A	p.(Val90Met)	N (0.48)	Neutral (0.166, 7)	DELETERIOUS (0.55)	No
VHL	1	c.242C>T	p.(Pro81Leu)	D (0.64)	Neutral (0.378, 2)	DELETERIOUS (0.51)	No
VHL	1	c.338G>A	p.(Arg113Gln)	D (0.62)	Neutral (0.198, 6)	NEUTRAL (0.83)	No
VHL	2	c.361G>C	p.(Asp121His)	D (0.67)	Disease (0.73, 5)	DELETERIOUS (0.87)	Yes (Pathogenic)
VHL	3	c.554A>G	p.(Tyr185Cys)	D (0.64)	Disease (0.605, 2)	DELETERIOUS (0.76)	Yes (Pathogenic)
VHL	3	c.629G>A	p.(Arg210Gln)	D (0.57)	Neutral (0.168, 7)	NEUTRAL (0.75)	No

Table 5.30 Concordance between types of *in silico* prediction tools

Red text indicates the variant is considered to be damaging, yellow indicates the effect is unclear and green indicates the variant is predicted to be benign. Concordance between all sub-groups of program is demonstrated in the last column with subgroups without concordance excluded.

				Sequence and evolutionary conservation	Protein sequence and structure	Machine learning	Meta-tools	
Gene	**Exon**	**Variant (codingDNA)**	**Variant (protein)**	**≥75% concordant**	**Polyphen-2 Concordant**	**≥75% Concordant**	**100% Concordant**	**SUBGROUPS CONCORDANT?**
FLCN	7	c.715C>T	p.(Arg239Cys)	Yes (Pathogenic)	Yes (Pathogenic)	No	Yes (Pathogenic)	**Yes (Pathogenic)**
FLCN	12	c.1333G>A	p.(Ala445Thr)	Yes (Benign)	Yes (Benign)	No	Yes (Benign)	**Yes (Benign)**
MAX	5	c.425C>T	p.(Ser142Leu)	Yes (Benign)	Yes (Pathogenic)	No	No	**No**
SDHA	2	c.136A>G	p.(Lys46Glu)	Yes (Benign)	Yes (Benign)	No	No	**Yes (Benign)**
SDHA	8	c.923C>T	p.(Thr308Met)	Yes (Pathogenic)	Yes (Pathogenic)	Yes (Pathogenic)	Yes (Pathogenic)	**Yes (Pathogenic)**
SDHA	10	c.1273G>A	p.(Val425Met)	Yes (Benign)	Yes (Poss. path)	Yes (Benign)	Yes (Benign)	**No**
SDHAF2	3	c.319C>T	p.(Arg107Cys)	Yes (Pathogenic)	No	No	No	**No**
SDHB	1	c.32G>A	p.(Arg11His)	Yes (Benign)	No	No	No	**No**
SDHB	5	c.487T>C	p.(Ser163Pro)	Yes (Benign)	Yes (Benign)	No	Yes (Benign)	**Yes (Benign)**
SDHC	3	c.148C>T	p.(Arg50Cys)	Yes (Pathogenic)	Yes (Pathogenic)	Yes (Pathogenic)	Yes (Pathogenic)	**Yes (Pathogenic)**
SDHC	4	c.214C>T	p.(Arg72Cys)	Yes (Pathogenic)	No	Yes (Pathogenic)	Yes (Pathogenic)	**No**
SDHC	5	c.380A>G	p.(His127Arg)	Yes (Pathogenic)	Yes (Pathogenic)	Yes (Pathogenic)	Yes (Pathogenic)	**Yes (Pathogenic)**
SDHD	1	c.34G>A	p.(Gly12Ser)	Yes (Benign)	Yes (Benign)	No	Yes (Benign)	**Yes (Benign)**
TMEM127	3	c.268G>A	p.(Val90Met)	Yes (Benign)	Yes (Poss. path)	No	No	**No**
VHL	1	c.242C>T	p.(Pro81Leu)	No	No	No	No	**No**
VHL	1	c.338G>A	p.(Arg113Gln)	Yes (Benign)	No	No	No	**No**
VHL	2	c.361G>C	p.(Asp121His)	Yes (Pathogenic)	Yes (Pathogenic)	Yes (Pathogenic)	Yes (Pathogenic)	**Yes (Pathogenic)**
VHL	3	c.554A>G	p.(Tyr185Cys)	No	Yes (Benign)	Yes (Pathogenic)	Yes (Pathogenic)	**No**
VHL	3	c.629G>A	p.(Arg210Gln)	Yes (Benign)	Yes (Benign)	No	No	**Yes (Benign)**

Table 5.31 Comparison of concordance between amalgamated results of all programs and the amalgamated results of the three programs used by diagnostic laboratories

Gene	Exon	Variant (coding DNA)	Variant (protein)	Amalgamation of results for all tools (n=19) From Table 5.30	Amalgamated diagnostic tools (n=3) - 100% concordant only From Table 5.26
FLCN	7	c.715C>T	p.(Arg239Cys)	Yes (Pathogenic)	YES (Pathogenic)
FLCN	12	c.1333G>A	p.(Ala445Thr)	Yes (Benign)	YES (Benign)
MAX	5	c.425C>T	p.(Ser142Leu)	NO	NO
SDHA	2	c.136A>G	p.(Lys46Glu)	Yes (Benign)	YES (Benign)
SDHA	8	c.923C>T	p.(Thr308Met)	Yes (Pathogenic)	YES (Pathogenic)
SDHA	10	c.1273G>A	p.(Val425Met)	NO	NO
SDHAF2	3	c.319C>T	p.(Arg107Cys)	NO	NO
SDHB	1	c.32G>A	p.(Arg11His)	NO	NO
SDHB	5	c.487T>C	p.(Ser163Pro)	Yes (Benign)	YES (Benign)
SDHC	3	c.148C>T	p.(Arg50Cys)	Yes (Pathogenic)	NO
SDHC	4	c.214C>T	p.(Arg72Cys)	NO	NO
SDHC	5	c.380A>G	p.(His127Arg)	Yes (Pathogenic)	NO
SDHD	1	c.34G>A	p.(Gly12Ser)	Yes (Benign)	YES (Benign)
TMEM127	3	c.268G>A	p.(Val90Met)	NO	NO
VHL	1	c.242C>T	p.(Pro81Leu)	NO	NO
VHL	1	c.338G>A	p.(Arg113Gln)	NO	NO
VHL	2	c.361G>C	p.(Asp121His)	Yes (Pathogenic)	NO
VHL	3	c.554A>G	p.(Tyr185Cys)	NO	NO
VHL	3	c.629G>A	p.(Arg210Gln)	Yes (Benign)	YES (Benign)

5.4.4 RNA analysis for VHL

Analysis of the effect variants may have upon the correct splicing of *VHL* mRNA is complicated by the isoforms normally produced. VHL is a small gene consisting of only three exons. There are two mRNA isoforms: 1) NM_000551, which includes all 3 exons (designated here as '*VHL* full') and 2) NM_198156, which lacks exon 2, an inframe coding exon (designated here as '*VHL* Δ2') (135-137). Given that exon 2 deletions are pathogenic mutations leading to VHL type 1, the isoform lacking exon 2 cannot be sufficient for VHL protein function. Mutations that affect the consensus splice sites of exons 1, 2 and 3, again causing skipping of exon 2, are also known to be pathogenic (137;144). Analysis of the RNA is further hindered by the paucity of coding polymorphic variants, which does not allow discernment of allele-specific expression.

Fresh blood samples for RNA extraction were obtained from a patient with a clinical diagnosis of VHL with an intronic deletion of 5 base pairs, c.[341-21_341-17delAACCT];[=]. *In silico* prediction software only showed minimally reduced scores as a result of this variant (see Figure 5.6). Full gene sequencing was carried out (see Chapter 3) but no additional likely causative variants were identified. Given the phenotype, and the fact that the intronic deletion had not been reported on any of the variation databases, it was felt that further analysis was warranted. The potential splicing effects that this variant could have were postulated in section 5.3.3.1 .

A number of assays were designed that could be used on this patient and other patients with potential splicing mutations in the VHL gene. Initially a set of assays combining RT-PCR,

gel electrophoresis and Sanger sequencing were designed and performed, followed by an assay using droplet digital RT-PCR.

5.4.4.1 VHL RNA analysis using RT-PCR, gel electrophoresis and Sanger sequencing

The first assay was designed to amplify cDNA for both *VHL* isoforms. All cDNAs analysed showed bands at the expected sizes for the two isoforms, which differ in size by 123 bp (see Figure 5.8 and Figure 5.9). The intensity of the upper band encoding '*VHL* full' appeared higher than that of the '*VHL* Δ2' isoform band for the 4 control cDNAs, despite variability in amplification efficiency. Meanwhile, the respective intensity of the bands appeared reversed for the *VHL* c.[341-21_341-17delAACCT];[=] patient's cDNA bands (see Figure 5.9). This was confirmed upon Sanger sequencing because the relative abundance of the two isoforms could be observed on the electropherograms (see Figure 5.10). The '*VHL* full' was the predominant isoform in the normal control samples and the '*VHL* Δ2 'isoform was predominant in the cDNA from the patient with the intronic 5 bp deletion.

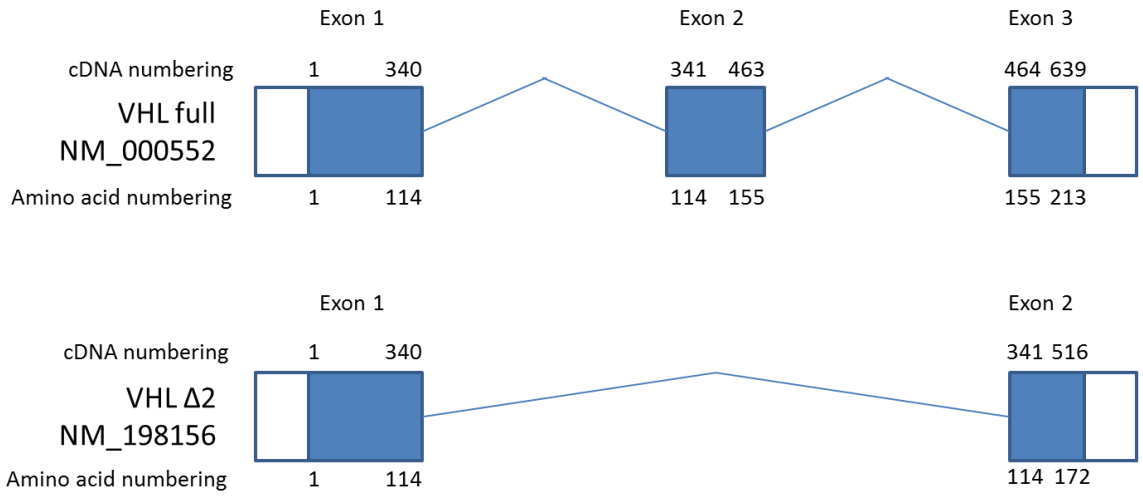


Figure 5.8 Schematic comparing the two transcribed isoforms of *VHL*
'*VHL* full' (NM_000552) which encodes the functional protein and '*VHL* Δ2' (NM_198156) which lacks exon two and doesn't produce a functional protein.

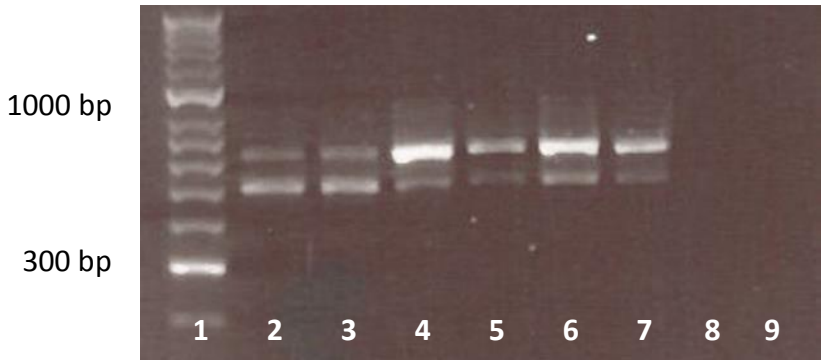


Figure 5.9 Amplification of coding region of *VHL* from cDNA to include both isoforms
Lanes: 1) Hyperladder II; 2&3) c.[341-21_341-17delAACCT];[=] cDNAs; 4-7) Normal control cDNAs; 8) Negative control from cDNA preparation; 9) Assay water control.

a)

Full ATGCCCCGGAGGGCGGAGAACTGGGACGAGGCCGAGGTAGGCGCGGAGGAGGCAGGCGTC

Δ2 ATGCCCCGGAGGGCGGAGAACTGGGACGAGGCCGAGGTAGGCGCGGAGGAGGCAGGCGTC

Full GAAGAGTACGGCCCTGAAGAAGACGGCGGGGAGGAGTCGGGCGCCGAGGAGTCCGGCCCG

Δ2 GAAGAGTACGGCCCTGAAGAAGACGGCGGGGAGGAGTCGGGCGCCGAGGAGTCCGGCCCG

Full GAAGAGTCCGGCCCGGAGGAACTGGGCGCCGAGGAGGAGATGGAGGCCGGGCGGCCGCGG

Δ2 GAAGAGTCCGGCCCGGAGGAACTGGGCGCCGAGGAGGAGATGGAGGCCGGGCGGCCGCGG

Full CCCGTGCTGCGCTCGGTGAACTCGCGCGAGCCCTCCCAGGTCATCTTCTGCAATCGCAGT

Δ2 CCCGTGCTGCGCTCGGTGAACTCGCGCGAGCCCTCCCAGGTCATCTTCTGCAATCGCAGT

Full CCGCGCGTCGTGCTGCCCGTATGGCTCAACTTCGACGGCGAGCCGCAGCCCTACCCAACG

Δ2 CCGCGCGTCGTGCTGCCCGTATGGCTCAACTTCGACGGCGAGCCGCAGCCCTACCCAACG

Full CTGCCGCCTGGCACGGGCCGCCGCATCCACAGCTACCGAGGTCACCTTTGGCTCTTCAGA

Δ2 CTGCCGCCTGGCACGGGCCGCCGCATCCACAGCTACCGAGTGTATACTCTGAAAGAGCGA

Full GATGCAGGGACACACGATGGGCTTCTGGTTAACCAAACTGAATTATTTGTGCCATCTCTC

Δ2 TGCCTCCAGGTTGTCCGGAGCCTAGTCAAGCCTGAGAATTACAGGAGACTGGACATCGTC

Full AATGTTGACGGACAGCCTATTTTTGCCAATATCACACTGCCAGTGTATACTCTGAAAGAG

Δ2 AGGTCGCTCTACGAAGATCTGGAAGACCACCCAAATGTGCAGAAAGACCTGGAGCGGCTG

Full CGATGCCTCCAGGTTGTCCGGAGCCTAGTCAAGCCTGAGAATTACAGGAGACTGGACATC

Δ2 ACACAGGAGCGCATTGCACATCAACGGATGGGAGATTGA

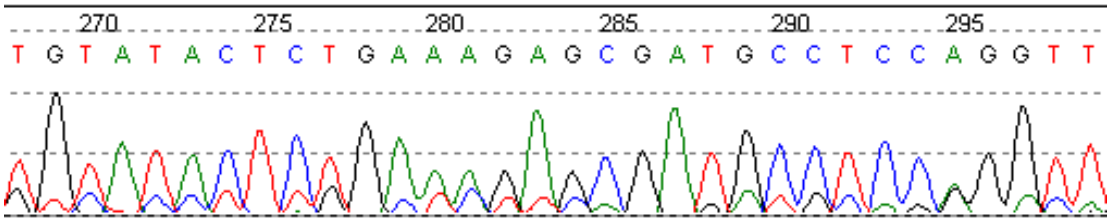
Full GTCAGGTCGCTCTACGAAGATCTGGAAGACCACCCAAATGTGCAGAAAGACCTGGAGCGG

Δ2

Full CTGACACAGGAGCGCATTGCACATCAACGGATGGGAGATTGA.....

Δ2

b)



c)

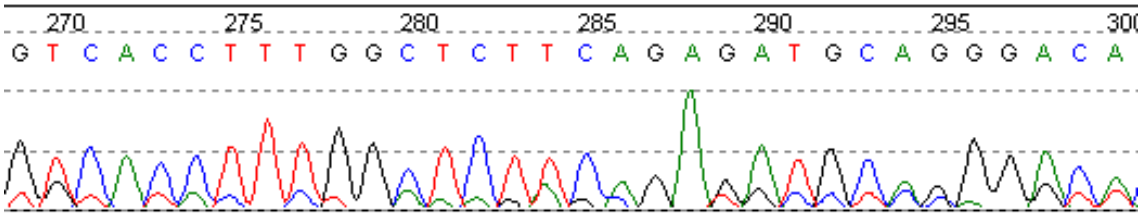


Figure 5.10 Expected sequence and Sanger sequencing of amplification product of coding region of *VHL* to include both isoforms

a) Alignment of the two cDNA isoforms to show the expected sequence in the forward direction following Sanger sequencing, exon 1 in purple, exon 2 in blue and exon 3 in green; b) Example of electropherogram of Sanger sequencing of amplification of c.[341-21_341-17delAACCT];[=] cDNA, commencing after the last base of exon 1; c) Example electropherogram of Sanger sequencing of a normal control.

The second assay was designed to amplify only the *VHL* Δ2 isoform by placing a primer over the junction of exon 1 and exon 3. As expected, all cDNAs amplified the *VHL* Δ2 isoform (see Figure 5.11). During the full *VHL* gene NGS analysis performed in Chapter 4, a common Class I polymorphism was identified heterozygously in the 3'UTR, c.[*294G>A];[=] for the patient with the 5 bp deletion. Sanger sequencing was used to determine if both alleles were found in the mRNA by examining this polymorphic base. Figure 5.12 shows that the polymorphism was present heterozygously, indicating both alleles were present.

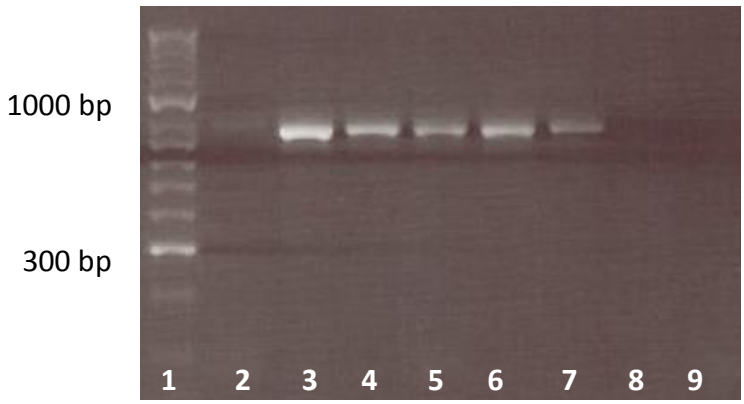


Figure 5.11 Amplification of the coding region of *VHL* from cDNA for only the VHL Δ2 isoform
Lanes: 1) Hyperladder II; 2&3) c.[341-21_341-17delAACCT];[=] cDNAs; 4-7) Normal control cDNAs; 8) Negative control from cDNA preparation; 9) Assay water control.

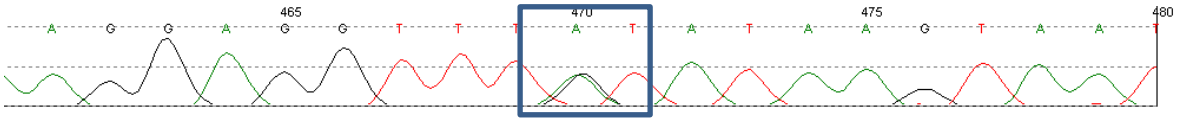


Figure 5.12 Sanger sequencing of cDNA for *VHL* Δ2 isoform only for intronic deletion patient
The c.*294G>A polymorphism, marked with box, is identified heterozygously in the cDNA.

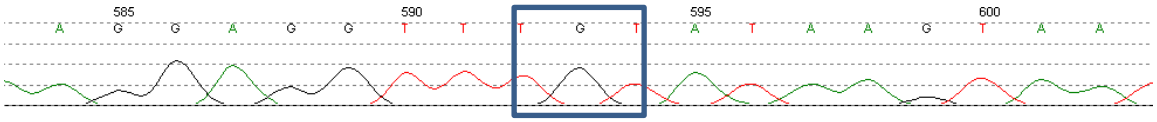


Figure 5.13 Sanger sequencing of cDNA for *VHL* full isoform only for intronic deletion patient
The c.*294G>A polymorphism, marked with a box, is identified homozygously in the cDNA.

The third assay was a replica of the second, but the primer was designed to bind over the exon 1 to exon 2 junction, so only the '*VHL* full' isoform would be amplified. Again all

samples produced an amplicon but upon sequencing of the cDNA, the polymorphism was observed homozygously, indicating only one allele was present and thus that no mRNA was present from one allele for the '*VHL* full' isoform.

5.4.4.2 VHL RNA analysis using droplet digital RT-PCR

A droplet digital RT-PCR assay was designed to confirm semi-quantitatively that the two isoforms were abnormally represented in the cDNA of the patient with c.[341-21_341-17delAACCT];[=], designated patient A for this assay.

Primers and probes were designed to amplify and specifically detect each of the two isoforms. The concentration (copies/µl) of each isoform was determined for the patient sample and eight normal control cDNA samples. Then the relative abundance of the '*VHL* full' isoform was compared to the '*VHL* Δ2' isoform for each sample. This was performed in duplicate and the average relative abundance plotted (see Figure 5.14). The difference between the patient data and the normal control data was considered extremely statistically significant using the unpaired t-test giving a two-tailed P value of less than 0.0001.

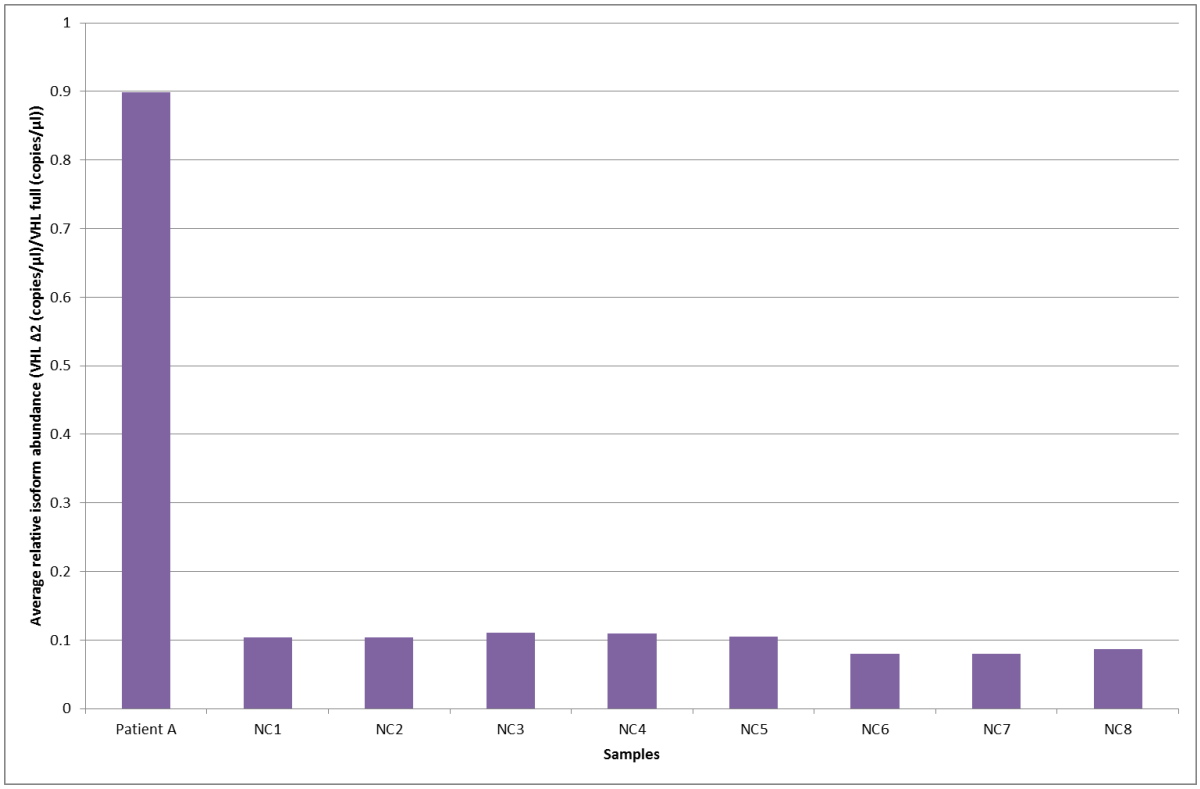


Figure 5.14 Graph showing the average relative abundance of the *VHL* Δ2 isoform compared to the *VHL* full isoform for the intronic deletion patient compared to eight normal controls

A very limited number of additional RNAs were also available from other patients who had been analysed using the *VHL* NGS full gene assay. These five cDNAs were also input into the assay and the statistical significance of the results considered (see Table 5.32 and Figure 5.15). While the difference in the average relative abundance of the isoforms for Patient C was considered statistically significant, this was not the case for any of the other patients as compared with normal controls. *VHL* c.[463+8C>T];[=] was not predicted in silico to affect the normal splicing of the *VHL* gene either by decreasing the score of the true donor site or creation of a cryptic splice site, however the result was of interest given the variant's location within the splice donor consensus sequence. Analysis of the RNA by PCR and sequencing did not identify an aberrant splicing event, however there were no heterozygous

variants in the sequence analysed and therefore loss of an aberrant transcript by NMD cannot be discounted (RT-PCR and sequence analysis of cDNA for this patient performed by WMRGL).

Table 5.32 Statistical significance of difference of relative abundance of isoforms for six patients examined using the digital droplet PCR method

Patient	*VHL* variant identified	Predicted to affect normal splicing of VHL exon 2	Statistically significant isoform difference
A	c.[341-21_341-17delAACCT];[=]	YES	YES (P = >0.0001)
B	c.[463+8C>T];[=]	NO	NO
C	c.[394C=/394C>T];[=] p.(Gln132Ter)[a]	YES[b]	YES (P = >0.0001)
D	c.[*3082C>T];[=]	NO	NO
E	No mutation identified	NO	NO
F	No mutation identified	NO	NO

[a] Variant observed at a mosaic level of approximately 7% in lymphocyte DNA. [b] This mutation sits within exon 2, it creates a stop codon at amino acid 132 and the transcript would be expected to undergo nonsense mediated decay (NMD).

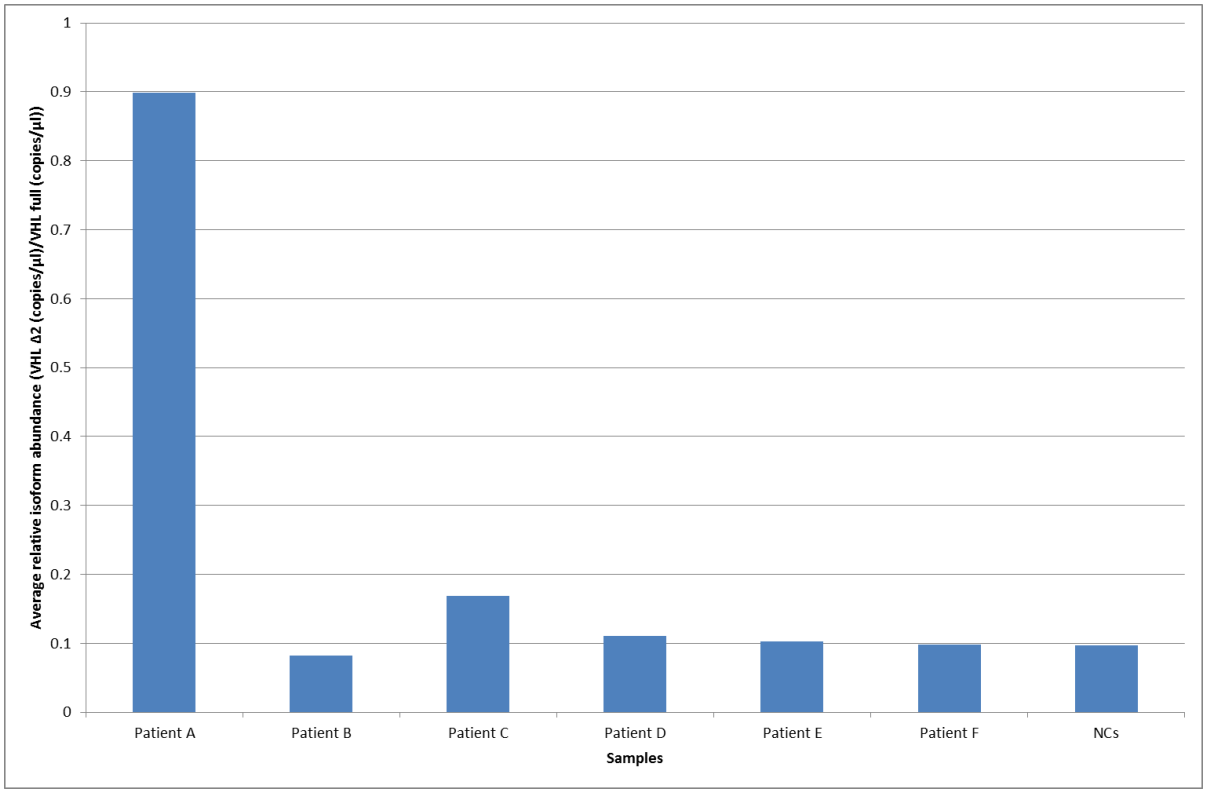


Figure 5.15 Graph showing the average relative abundance of the '*VHL* Δ2' isoform compared to the '*VHL* full' isoform for six patients compared to the average of eight normal controls

5.4.5 FFPE tumour analysis

Analysis of FFPE tumour material from patients with a number of variants was performed. Unfortunately tumour material was not available for the majority of patients in whom the Class II and III variants had been identified. However it was available for some and additional tumour material was available from some patients with variants which had been identified previously by the WMRGL. For all variants where tumour material was available, loss of heterozygosity analysis was performed by direct Sanger sequencing of the tumour tissue and comparison with the Sanger sequencing for the lymphocyte DNA from that patient. Additionally, if it was possible to do so, immunohistochemistry (IHC) was performed for any variant in *SDHB*, *SDHC* and *SDHD* using immunohistochemistry for SDHB. The IHC was performed in collaboration with the Department of Molecular Pathology, University Hospital Birmingham NHS Trust.

Table 5.33 Summary of direct tumour analysis using loss of heterozygosity (LOH) analysis and immunohistochemistry

In silico prediction is the consensus of the three diagnostically used splicing algorithms. The results for patient 1 can be seen in Figure 5.19.

Patient	Gene	coding DNA	Protein	LOH	IHC result	In silico prediction	Identified by
1	*SDHB*	c.220G>A	p.(Asp74Asn)	Yes - Partial loss of WT allele	-	Discordant	Other research
2	*SDHB*	c.269G>A	p.(Arg90Gln)	Yes - Total loss of WT allele	SDHB - Patchy (? Fixation artefact)	Discordant	WMRGL
3	*SDHB*	c.296G>A	p.(Gly99Asp)	Yes - Partial loss of WT allele	SDHB - Loss of staining	Pathogenic	WMRGL
4	*SDHB*	c.298T>C	p.(Ser100Pro)	Yes - Total loss of WT allele	-	Pathogenic	WMRGL
5	*SDHB*	c.326A>G	p.(Asn109Ser)	Yes - Total loss of WT allele	SDHB - Normal staining	Discordant	WMRGL
6	*SDHB*	c.487T>C	p.(Ser163Pro)	No	-	Benign	WMRGL/Thesis/ Other research
				No	SDHB - Normal staining		
				No	SDHB - Normal staining		
				No	SDHB - Normal staining		
				Yes - Partial loss of variant allele	-		
				No	-		
				No	-		
7	*SDHB*	c.529C>G	p.(Arg177Gly)	Yes - Total loss of WT allele	-	Discordant	WMRGL
8*	*SDHB*	c.587G>A	p.(Cys196Tyr)	Yes - Partial loss of WT allele	SDHB - Weak diffuse staining	Pathogenic	WMRGL
9*	*SDHB*	c.590C>G	p.(Pro197Arg)	Yes - Total loss of WT allele	-	Pathogenic	WMRGL
10	*SDHB*	c.801G>T	p.(Lys267Asn)	Yes - Partial loss of WT allele	SDHB - Patchy (? Fixation artefact)	Pathogenic	WMRGL/Thesis
11	*SDHC*	c.148C>T	p.(Arg50Cys)	No	-	Discordant	WMRGL/Thesis
12	*SDHC*	c.380A>G	p.(His127Arg)	Failed	SDHB - Normal staining	Discordant	Thesis
13	*SDHD*	c.116C>T	p.(Pro39Leu)	No	SDHB - Normal staining	Benign	WMRGL
14	*SDHD*	c.34G>A	p.(Gly12Ser)	No	-	Benign	WMRGL/Thesis
				No	-		
				No	-		
15	*SDHD*	c.149A>G	p.(His50Arg)	No	-	Discordant	WMRGL

* Pathogenic mutations, included as seen in patient who also has variant

5.4.6 Functional analysis of VHL protein stability

A simple functional assay was designed to investigate the stability of VHL proteins when the *VHL* gene had been mutated. A similar assay had been used successfully in Prof E Maher's research group to investigate the stability of the folliculin protein when the *FLCN* gene had been mutated (248). The VHL assay comprised modelling the *VHL* variants in a pcDNA3.1 plasmid containing wild type HA tagged VHL using site directed mutagenesis, transforming them into a bicistronic vector which also expressed green fluorescent protein (GFP) and transiently transfecting them into HEK293T cells. The level of VHL protein expression was measured using densitometry of western blotted protein lysate by dividing the amount of HA, as a measurement of VHL protein expression, by GFP protein, as an indicator of the level of transfection for each variant.

5.4.6.1 Previous work

Site directed mutagenesis and transformation had already been performed by Prof Maher's group for several *VHL* sequence variants. The variants were in the pcDNA3.1 plasmid as human influenza hemagglutinin (HA) tagged-*VHL;* the cDNA included codons 1-53 which allowed translation of both protein isoforms. These variants included examples of various different classes of *VHL* pathogenic mutations which could be used as controls in the functional experiments (see Table 5.34).

Table 5.34 List of available VHL variants to use as controls

Variant	Variant type	VHL type
c.74C>T; p.(Pro25Leu)	Missense	Polymorphism
c.194C>G; p.(Ser65Trp)	Missense	VHL type 1
c.250G>T; p.(Val84Leu)	Missense	VHL type 2C
c.361_362insC; p.(Asp121Alafs*11)	Premature truncation	VHL type 1
c.499G>C; p.Arg167Gly	Missense	VHL type 2B/C
c.562C>G; p.Leu188Val	Missense	VHL type 2C (few reports of RCC)

5.4.6.2 Additional variant selection

A number of additional VHL variants were selected to be included in the assay. These compromised VUSs identified during routing diagnostic testing by the West Midlands Regional Genetics Laboratory (WMRGL), some of which were also identified during this study (Table 5.35).

Table 5.35 VHL variants of unknown significance selected for inclusion in VHL protein stability assay

Gene	Exon	cDNA change	Protein change	Protein shorthand	Identified by
VHL	1	c.154G>A	p.(Glu52Lys)	E52K	WMRGL
VHL	1	c.242C>T	p.(Pro81Leu)	P81L	WMRGL/Thesis
VHL	1	c.245G>T	p.(Arg82Leu)	R82L	WMRGL
VHL	1	c.284C>G	p.(Pro95Arg)	P95R	WMRGL
VHL	1	c.326T>G	p.(Ile109Ser)	I109S	WMRGL
VHL	2	c.361G>C	p.(Asp121His)	D121H	WMRGL/Thesis
VHL	2	c.391A>C	p.(Asn131His)	N131H	WMRGL
VHL	3	c.605C>T	p.(Thr202Ile)	T202I	WMRGL

5.4.6.3 Site directed mutagenesis (SDM)

SDM was used to model the variants in full length VHL in pcDNA3.1. Following SDM, plasmid DNA was grown up and extracted by mini-prep from several individual colonies for each VUS. Then the complete coding region of VHL was Sanger sequenced to rule out the introduction of additional mutations during the mutagenesis process. This had occurred in several plasmids and ranged from additional single base changes to large deletions. However, 6/8 of the VUSs were modelled without additional mutations (see Figure 5.16). Two VUSs were not modelled despite repeated attempts. (see Figure 5.16).

Once the VUSs had been introduced into VHL by SDM, both the VUSs and the control mutation inserts were subcloned into the bicistronic vector, pIRES2-AcGFP1, by digestion,

gel extraction, ligation and transformation. Again the complete coding region of VHL was sequenced to rule out the introduction of mutations during this process.

5.4.6.4 Transient transfection and immunoblotting

All the constructs (VUSs and controls) were transiently transfected into HEK293T cells. The cells were then lysed and the protein extracted. The protein was visualised by immunoblotting using optimised antibody concentrations. The HA tag of VHL was probed as the cell line used produces pVHL protein. An example of the results achieved is shown in Figure 5.17.

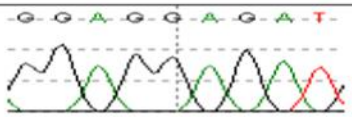
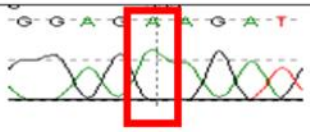
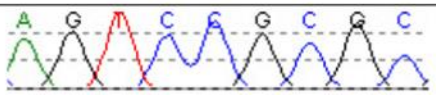
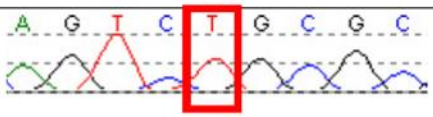
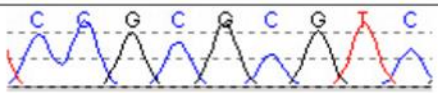
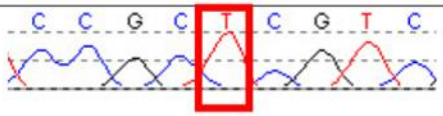
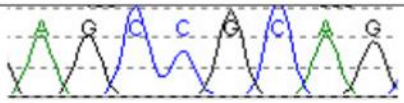
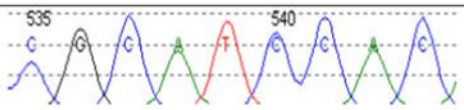
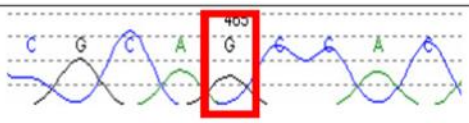
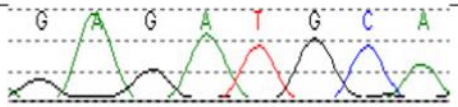
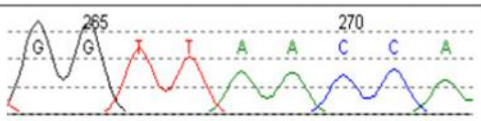
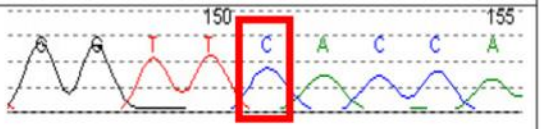
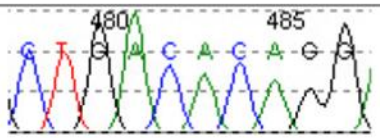
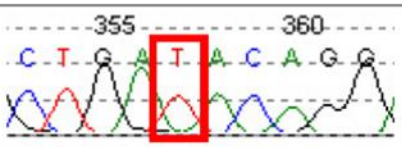
Variant of interest	Normal sequence	'Mutated' sequence
E52K		
P81L		
R82L		
P95R		Failed despite repeated attempts
I109S		
D121H		Failed despite repeated attempts
N131H		
T202I		

Figure 5.16 Results of site directed mutagenesis
The altered base is highlighted with a red box for each successful assay.

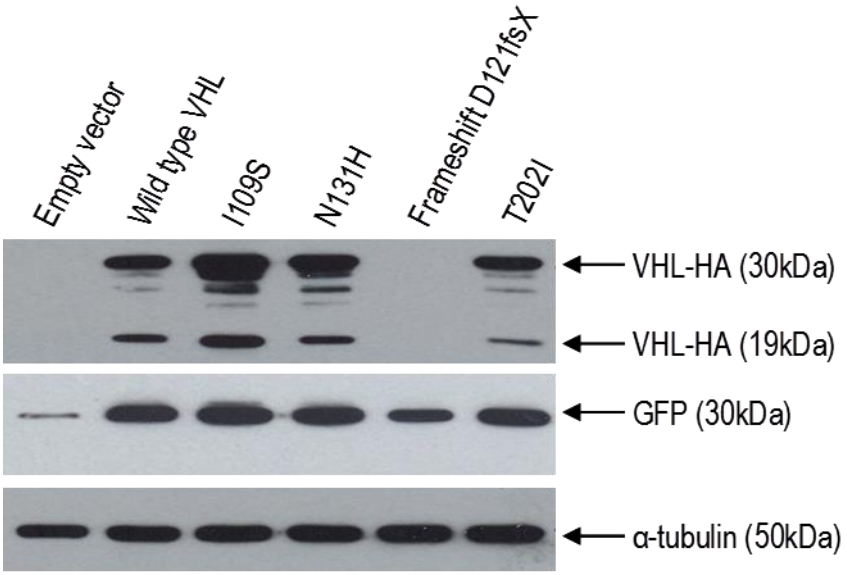


Figure 5.17 Example immunoblot for HA tag of the VHL construct, GFP and α-tubulin
VHL expression was variable across the 3 variants examined here. However the variation was not reproducible between experiments. The controls of the empty vector and the frameshift both behaved as expected with no pVHL-HA detected.

5.4.6.5 Dosimetry to compare stability of VUSs in VHL

The strength of each band was determined using densitometry (GeneTools, SyncGene). Each transfection was repeated 3 times and the average stabilities were calculated and plotted graphically. Unfortunately, as Figure 5.18 shows, VHL stability was not reproducible between experiments for the 19kDa isoform. The larger isoform also gave inconsistent results and therefore this was not considered to be a good method to functionally assess the pathogenicity of VHL VUSs.

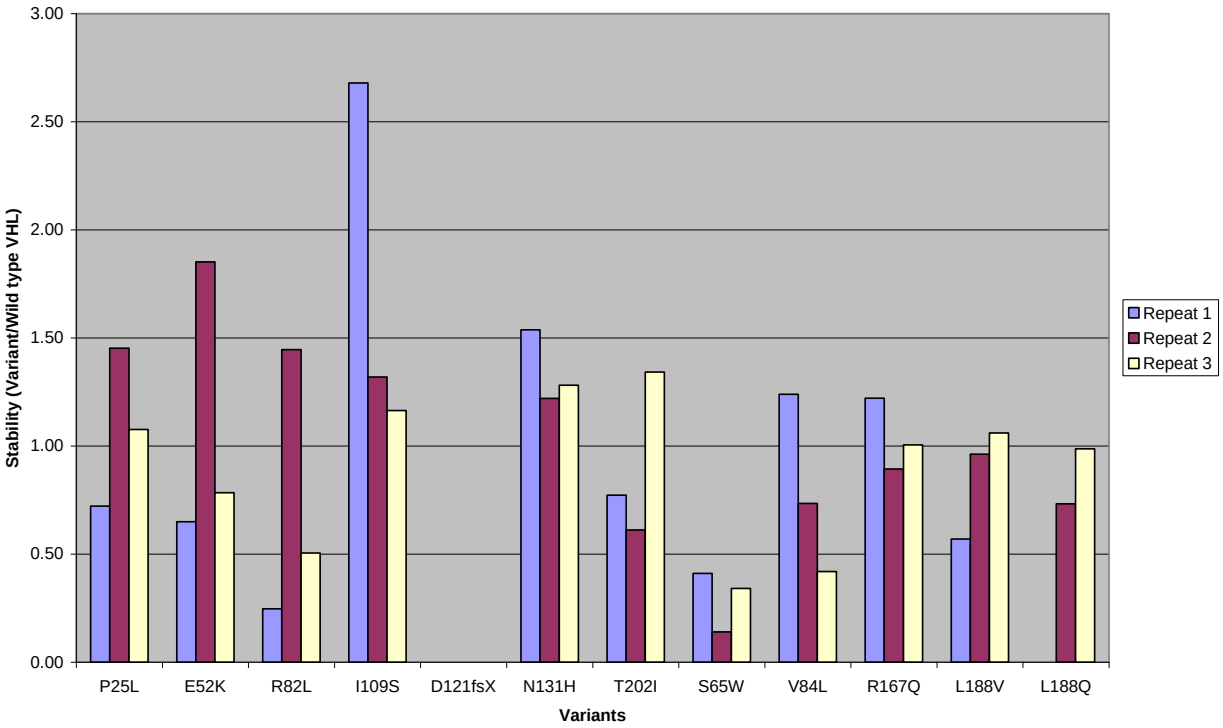


Figure 5.18 Graphical representation of the stability of variants in the 19kDa isoform of VHL-HA as a ratio of the stability of wildtype VHL-HA from three separate experiments
As expected, no HA was detected for the frameshift mutation as it would remove the HA tag. Stability was not reproducible between experiments.

5.4.7 Variant assessment summary

Fifty two variants were put through pathogenicity assessment using a number of laboratory based methods. 17.3% of these variants had their pathogenicity reclassified following these analyses. The majority of reclassifications resulted from additional variation data due to the release of the ExAC database, which allowed five variants to be reclassified as benign (I) and one as likely benign (II). One variant was classified as pathogenic following extensive RNA analysis. Two variants were reclassified as benign (I) following examination of multiple lines of evidence and are discussed in detail in section 5.6 .

Table 5.36 Summary of variant analysis and variants where their Class could be updated

Gene	Variant type	Variant (codingDNA)	Variant (protein)	Times observed	Original variant class	Seen at >1% in variation database	RNA analysis warranted?	RNA analysis performed	In silico missense predicted pathogenic (overall)	LOH	IHC (SDHB)	Updated variant class
FLCN	Missense	c.715C>T	p.(Arg239Cys)	2	VUS (III)	No	No		Pathogenic			No
FLCN	Missense	c.1333G>A	p.(Ala445Thr)	1	VUS (III)	No	No		Predicted benign			No
FLCN	3'UTR	c.*4A>G		1	VUS (III)	No	No		N/A			No
MAX	Missense	c.425C>T	p.(Ser142Leu)	1	VUS (III)	No	No		Discordant results			No
MET	Missense	c.3356G>C	p.(Gly1119Ala)	1	VUS (III)	No	No		N/A			No
SDHA	Missense	c.136A>G	p.(Lys46Glu)	1	VUS (III)	No	No		Predicted benign			No
SDHA	Synonymous	c.549C>T	p.(=)	1	VUS (III)	Yes						Benign (I)
SDHA	Synonymous	c.822C>T	p.(=)	1	VUS (III)	Yes						Benign (I)
SDHA	Synonymous	c.1002G>A	p.(=)	1	VUS (III)	No	No		N/A			No
SDHA	Missense	c.923C>T	p.(Thr308Met)	1	VUS (III)	No	No		Pathogenic			No
SDHA	Missense	c.1273G>A	p.(Val425Met)	1	VUS (III)	No	No		Discordant results			No
SDHA	Frameshift	c.1338delA	p.(His447Metfs*23)	1	VUS (III)	No	No		N/A			No
SDHA	Synonymous	c.1623G>A	p.(=)	1	VUS (III)	No	No		N/A			No
SDHA	Synonymous	c.1776T>C	p.(=)	1	VUS (III)	Yes						Benign (I)
SDHAF2	Missense	c.319C>T	p.(Arg107Cys)	1	VUS (III)	No	No		Discordant results			No
SDHB	Missense	c.32G>A	p.(Arg11His)	1	VUS (III)	No	No		Discordant results			No
SDHB	Missense	c.487T>C	p.(Ser163Pro)	6	VUS (III)	Yes*	No		Predicted benign	No LOH	No loss	Benign (I)*
SDHC	5'UTR	c.-118_-117delAG		1	VUS (III)	No	No		N/A			No
SDHC	Missense	c.148C>T	p.(Arg50Cys)	1	VUS (III)	No	No		Pathogenic	No LOH		No
SDHC	Missense	c.214C>T	p.(Arg72Cys)	1	VUS (III)	No	No		Discordant results			No
SDHC	Missense	c.380A>G	p.(His127Arg)	1	VUS (III)	No	No		Pathogenic	Failed	No loss	No
SDHD	Missense	c.34G>A	p.(Gly12Ser)	3	VUS (III)	Yes*	No		Predicted benign	No LOH	No loss	Benign (I)*
SDHD	Synonymous	c.312C>T	p.(=)	1	VUS (III)	Yes						No

Gene	Variant type	Variant (codingDNA)	Variant (protein)	Times observed	Original variant class	Seen at >1% in variation database	RNA analysis warranted?	RNA analysis performed	In silico missense predicted pathogenic (overall)	LOH	IHC (SDHB)	Updated variant class
TMEM127	Missense	c.268G>A	p.(Val90Met)	1	VUS (III)	No	No		Discordant results			No
TMEM127	Synonymous	c.411T>A	p.(=)	1	VUS (III)	No	No		N/A			No
TMEM127	Synonymous	c.534C>T	p.(=)	1	VUS (III)	No**	Yes	Not available	N/A			Likely benign (II)No
VHL	5'UTR	c.-3933C>T		2	VUS (III)	No	No		N/A			No
VHL	5'UTR	c.-3197_-3195delCTC		1	VUS (III)	No	No		N/A			No
VHL	5'UTR	c.-963G>A		1	VUS (III)	No	No		N/A			No
VHL	Synonymous	c.183C>G	p.(Pro61Pro)	1	VUS (III)	No	No		N/A			No
VHL	Missense	c.242C>T	p.(Pro81Leu)	1	VUS (III)	No	No		Discordant results			No
VHL	Missense	c.338G>A	p.(Arg113Gln)	1	VUS (III)	No	No		Discordant results			No
VHL	Intronic	c.340+203G>A		1	VUS (III)	No	No		N/A			No
VHL	Intronic	c.340+280T>G		1	Likely benign (II)	No	No		N/A			No
VHL	Intronic	c.340+376C>A		1	Likely benign (II)	No	No		N/A			No
VHL	Intronic	c.341-123G>T		1	VUS (III)	No	No		N/A			No
VHL	Intronic	c.341-21_341-17delAACCT		1	VUS (III)	No	Yes	Aberrant splicing	N/A			Pathogenic (V)
VHL	Missense	c.361G>C	p.(Asp121His)	1	VUS (III)	No	No		Pathogenic			No
VHL	Intronic	c.463+8C>T		1	VUS (III)	No	No	Normal splicing	N/A			Likely benign (II)
VHL	Intronic	c.464-10G>A		1	VUS (III)	No	No		N/A			No
VHL	Intronic	c.464-1212A>G		1	Likely benign (II)	Yes						Benign (I)
VHL	Intronic	c.464-1434C>T		2	Likely benign (II)	No	No		N/A			No
VHL	Intronic	c.464-1530C>T		1	VUS (III)	No	No		N/A			No
VHL	Missense	c.554A>G	p.(Tyr185Cys)	1	VUS (III)	No	No		Discordant results			No
VHL	Missense	c.629G>A	p.(Arg210Gln)	1	VUS (III)	No	No		Predicted benign			No

Gene	Variant type	Variant (codingDNA)	Variant (protein)	Times observed	Original variant class	Seen at >1% in variation database	RNA analysis warranted?	RNA analysis performed	In silico missense predicted pathogenic (overall)	LOH	IHC (SDHB)	Updated variant class
VHL	Synonymous	c.639T>C	p.(Asp213Asp)	1	VUS (III)	No	No		N/A			No
VHL	3'UTR	c.*820A>G		1	Likely benign (II)	No	No		N/A			No
VHL	3'UTR	c.*2511A>G		1	VUS (III)	No	No		N/A			No
VHL	3'UTR	c.*3021T>C		1	VUS (III)	No	No		N/A			No
VHL	3'UTR	c.*3082C>T		1	VUS (III)	No	No		N/A			No
VHL	3'UTR	c.*3170G>A		1	VUS (III)	No	No		N/A			No
VHL	3'UTR	c.*3482dupA		1	Likely benign (II)	No	No		N/A			No

* There is further discussion regarding the pathogenicity of these variants in section 5.6 ;**Only reaches 1% population frequency in the GBR subpopulation.

5.5 Discussion

During the course of the gene screening performed as part of this research, 72 heterozygous variants were identified, excluding Class I benign polymorphisms. An additional three variants were detected at a mosaic level. Of these variants, 20/72 heterozygous variants and 3/3 mosaic level variants were considered to be Class IV - Likely pathogenic or Class V - Pathogenic and therefore clinically actionable. The remaining 52 variants were considered to be Class III VUS or Class II likely benign.

Of the variants listed there are some that sit outside standard practice, for example those variants identified in SDHA, discussed below, and those variants at ≥1% frequency in the population but considered in the literature to be putatively disease causing (discussed in section 5.6).

Multiple variant assessment techniques were used for the 52 variants of unclear significance. The first technique was reanalysis of the variants using the most up to date variation databases. This was the most successful method of variant reclassification. Any variant that was in ≥1% of alleles in any 'super population' was determined to be a Class I benign polymorphism. It could be argued that, given the incidence of disease related to all of the genes included is less than 1 in 10, 000, this threshold could be lowered further. However because at least one of the databases used is known to contain a cancer cohort 1% was considered appropriately cautious.

Next *in silico* splicing analysis was performed. Whilst it is recognised that *in silico* analysis can never be used to assign pathogenicity, it is a very helpful indicator of where RNA analysis is worthwhile. In this cohort of patients all predictions of altered splicing were

subtle and comparison of predicted cryptic splice site scores to the scores of the true splice sites for a given gene was a useful method of establishing if splicing analysis might be worthwhile. Only two variants were deemed to be worthwhile candidates for direct RNA analysis and RNA could only be obtained from one of these, *VHL* c.341-21_341-17delAACCT.

RNA analysis allowed reclassification of *VHL* c.341-21_341-17delAACCT from Class III - VUS to Class V - Pathogenic. This was not a simple task as it was predicted to cause skipping of exon 2 of *VHL,* which is also a known alternative mRNA isoform of *VHL*. Thus it was necessary to demonstrate that exon 2 skipping was occurring in a pathogenic manner. It was shown, through RT-PCR and sequencing, that only one allele was producing any of the functional full *VHL* mRNA isoform and, through semi-quantitative PCR, that the level of the *VHL* mRNA lacking exon two was disproportionate to level of normally spliced *VHL* mRNA, when compared to normal controls.

The semi-quantitative assay was also used to analyse three other VHL variants, from which patient RNA could be obtained. Neither *VHL* c.463+8C>T nor *VHL* c.*3082C>T showed any evidence of abnormal levels of exon 2 skipping. This is especially interesting for the intronic variant because, combined with the evidence of lack an alternatively splice product, it confirms the *in silico* finding that splicing would not be affected. The only caveat is that the presence of an aberrant mRNA, that has undergone NMD, cannot be discounted by this method. *VHL* c.[394C=/394C>T] p.(Gln132Ter) was considered to show a statistically significant difference (P = >0.0001) in the proportion of the two isoforms, with the non-functional isoform lacking exon two being overrepresented. This is likely to be the result of NMD of the nonsense transcript (since the nonsense mutation sits in exon 2, only transcripts

containing exon 2 can be affected). The proportional difference in transcript abundance is smaller as the mutation is present at a mosaic level in the blood, shown to be in 7% of reads by NGS rather than 50% for a heterozygous change, such as *VHL* c.341-21_341-17delAACCT.

In silico missense variant analysis was also performed for 19 variants in TSGs. Again, it is recognised that *in silico* analysis can never be used to definitively assign pathogenicity but rather to identify likely benign or pathogenic variants and distinguish those where further investigation is most warranted. The programs that used sequence and evolutionary conservation to predict pathogenicity were most concordant (58% totally concordant). The machine learning tools were least concordant (21% totally concordant), which is likely due to the different data sets used to train the algorithms. The results of the different program categories were combined and used to assign an overall prediction, since it was considered that the more sources used to predict pathogenicity where the data concurred (≥75% concordance within a category) the more accurate the prediction was likely to be. Finally the amalgamated predictions of the 17 programs were compared to the predictions using only the three tools routinely used in diagnostic laboratories. Interestingly there were no discrepant predictions. There were three additional predictions for the amalgamated set, likely to be the result of the more permissive concordance threshold set. These data suggest that the three diagnostically used prediction tools provide similar results to those produced using a much greater number of programs but their use is vastly less time consuming. Given that in *silico* analysis alone cannot be used for pathogenicity prediction, this work only enabled a group of variants to be identified for which additional pathogenicity investigations may be more or less fruitful.

Tumour analysis was performed for a number of variants. Not all variants analysed were those identified during this research since additional VUS were included that had been identified during previous Sanger analysis by the WMRGL. Interestingly, loss of heterozygosity (LOH) did not mean that loss of staining was observed by IHC analysis and these are conflicting results with regard to pathogenicity assessment. However, the mechanism of pathogenicity of a missense mutation in a TSG does not always have to be destabilisation of the protein (249), and so it is conceivable that LOH and retention of protein staining could coexist. Either LOH alone or LOH and loss of staining by IHC are supporting evidence for the pathogenicity of the applicable variants. Reassuringly, in all of the eight cases with concordant *in silico* predictions for a variant and LOH results, both results were congruent, five variants were predicted to be pathogenic and showed LOH with the variant allele retained and three cases were predicted to be benign and either no LOH was observed or LOH was observed with the variant allele lost.

In an attempt to functionally classify some of these variants, a protein stability assay for pVHL was attempted. Unfortunately, this failed to produce reproducible results. The attempt at this assay was time consuming and required specialised equipment not routinely available in the diagnostic setting, exemplifying why missense variant classification remains such a difficult procedure within the diagnostic laboratory.

The *SDHA* gene provided both analysis and variant interpretation challenges. It is a complex gene to analyse as it has 3 highly homologous pseudogenes. Additionally the clinical significance of the variants observed in the genes was not always clear cut. The c.91C>T p.(Arg31Ter) mutation was observed in 2/692 (0.3%) Dutch controls (73) and was reported

on the TCA Cycle Gene Mutation Database as 'Probably pathogenic' but 'probably very low penetrance' (250). Given that this is a nonsense mutation in the second exon of the gene its lack of penetrance casts a level of suspicion on the pathogenicity of this variant. Various analyses undertaken to prove the pathogenicity of this mutation are reported in the literature, such as identification of the second hit in the tumour. In the Kopershoek paper, *SDHA* c.91C>T p.(Arg31Ter) was identified in two individuals who both showed loss of the wild type allele in the tumours as the result of loss of heterozygosity (LOH) and immunohistochemistry (IHC) also showed loss of protein expression in the tumours (73). In addition, in the ExAC database (release October 2013), it was observed at an average MAF of only 0.02%, whilst Europeans demonstrated the highest MAF at 0.03% (18/66740) (207), much lower than the 0.3% in the original publication.

A frameshift mutation in *SDHA* was designated to be Class III in Chapter 3. At the time, it was considered that that the published evidence regarding *SDHA* pathogenicity was not conclusive enough to allow clinical decisions to be made on the basis of the finding. Therefore, it was felt that tumour studies to confirm loss of SDHA protein expression and identify the second hit should be performed prior to offering predictive testing to family members. However, during the course of this study the additional material to perform these analyses were not received.

5.6 Pathogenicity assessment of variants putatively associated with another independent cancer predisposition syndrome

There were two variants which were categorised as Class III - VUS, as a result of being putatively disease causing, despite having a known incidence in population studies of greater than 1%. These variants were *SDHB* c.487T>C p.(Ser163Pro) and *SDHD* c.34G>A p.(Gly12Ser). Both variants were classed as benign by the WMRGL, as a result of their population frequency, until a series of papers were published which associated them and other variants in the succinate dehydrogenase complex genes with an increased risk of Cowden syndrome, see Table 5.37 (246;247). As a result of these papers, further investigation of these variants has been undertaken to establish their pathogenicity.

The work performed comprises a brief literature review, examination of the frequency of identification of these variants during local diagnostic testing, and review of in silico and tumour analyses performed.

5.6.1 Literature review

Cowden syndrome (CS) is an autosomal dominant cancer predisposition syndrome, which is characterised by benign and malignant breast, thyroid and endometrial neoplasia, as well as macrocephaly and skin abnormalities (246). Up to 85% of patients have a germline mutation in *PTEN* (246). Cowden syndrome-like (CSL) describes individuals who have disease with Cowden syndrome features, but do not meet the strict International Cowden Consortium's criteria for the disease (247). Only about 5% of individuals with CSL have a germline mutation in *PTEN* (247).

In 2008, five variants in *SDHB* and *SDHD* were reported to be associated with the CS/CSL phenotype, see Table 5.37 (246). Functional studies using cell lysates from patient lymphoblastoid cell lines with these variants showed a combination of mitochondrial dysfunction exhibited by increased expression of manganese super oxide dismutase and/or reactive oxygen species (ROS) and activation of the MAPK and/or AKT pathways, which mimic PTEN dysfunction (246). The authors acknowledged that three of the variants (*SDHB* c.487T>C p.(Ser163Pro); *SDHD* c.34G>A p.(Gly12Ser); *SDHD* c.149A>G p.(His50Arg)) had been reported in specific population controls at between 1.1% and 3%, despite not being found in their 700 population controls, described as 'normal white populational controls of Northern and Western European origin'. Their explanation for this was that the individuals in the populations with known variant alleles may have had unrecognised or partial phenotypes (246). Notably, the tumour spectrum in the patients with an SDH variant was different to that of the patients with a *PTEN* mutation, with increased frequencies of breast, thyroid and renal cancers in the SDH variant set (246). The authors concluded that, whilst the data require validation in a larger study, *PTEN*-negative patients with CS/CSL should be offered SDH mutation analysis (246).

Table 5.37 Succinate dehydrogenase complex gene variants associated with Cowden disease

*variants identified in the course of this research in patients who are not known to have a Cowden syndrome like phenotype (Adapted from (247)).

Gene	codingDNA variation	Protein variation	Initially reported
SDHB	c.8C>G	p.(Ala3Gly)	2008 (246)
SDHB	c.9C>G	p.(Arg27Gly)	2012 (247)
SDHB	c.158G>A	p.(Gly53Glu)	2012 (247)
SDHB	c.170A>G	p.(His57Arg)	2012 (247)
SDHB	c.359A>G	p.(Asn120Ser)	2012 (247)
*SDHB**	c.487T>C	p.(Ser163Pro)	2008 (246)
SDHB	c.643G>A	p.(Ala215Thr)	2012 (247)
SDHC	c.197C>T	p.(Ala66Val)	2012 (247)
*SDHD**	c.34G>A	p.(Gly12Ser)	2008 (246)
SDHD	c.149A>G	p.(His50Arg)	2008 (246)
SDHD	c.433C>A	p.(His145Asn)	2008 (246)

In a letter to the editor of the American Journal of Human Genetics, Bayley examined the validity of some of the findings of the above paper (251). He questioned what normal controls had been used for the functional assays and stated that, given the frequency of some of the variants had been demonstrated to be as high as 1 in 20 in certain populations, the prevalence of the disease should be much higher in those populations (251). He commented that 'in light of the high frequency of these variants in many populations, adequate comprehension of their clinical significance and thus appropriate counselling will be very challenging' (251).

In 2012, Ni *et. al.* published a more extensive analysis of SDH complex gene variants in patients with CS/CSL (247). They identified variants in 8% (49/608) of *PTEN* mutation negative CS/CSL patients but also in 6% (26/444) of *PTEN* mutation positive CS/CSL patients, stating that an SDH variant modified *PTEN* cancer risks (247). This second study identified an additional five SDH variants, the majority of which had not been recorded previously on

dbSNP or in the SDH genes LSDBs but two had been reported in African American populations on dbSNP (247). Ni *et. al.* identified that their patient group had increased risk of breast cancer if they had both a *PTEN* mutation and an SDH variant compared to having either alone, and again found that SDH variants without *PTEN* mutation conferred an increased risk of breast cancer and epithelial thyroid cancer (247). The authors went on to perform various functional analyses on patient lymphoblastoid cell lines (247). They identified that SDH variant cell lines had increased cellular ROS, which was highest when there was also a *PTEN* mutation (247). SDH variant cell lines also showed stabilised HIF1α signalling and the majority showed loss of steady state p53 (247).

Interestingly, another paper was published in 2012 by Lendvai *et. al.* entitled 'Over-representation of the G12S polymorphism in patients with MEN2A syndrome'(252). This observed that 15.5% (8/55) patients with MEN2A had the *SDHD* c.34G>A p.(Gly12Ser) variant compared to 1% (1/100) of controls (252). Both the patients and controls were of Hungarian descent (252). No difference was detected in phenotype between the groups with and without the variant (252). Furthermore, upon review of the published literature a paper by Montani *et. al.* was identified where *SDHB* c.487T>C p.(Ser163Pro), *SDHD* c.34G>A p.(Gly12Ser) and *SDHD* c.149A>G p.(His50Arg) had been found more commonly in a population with medullary thyroid cancer (17.1% (6/35) combined) than in the control population (1.2% (1/83) combined) (253).

5.6.2 Local cohort frequency analysis

Of the list of variants implicated in CS/CSL, two were identified in NGS prospective patient analysis. In the PHEO/PGL/HNPGL cohort, *SDHB* c.487T>C p.(Ser163Pro) was identified

heterozygously in 7 of 239 patients, an allele frequency of 1.5% and SDHD c.34G>A p.(Gly12Ser) was identified heterozygously in 3 of 120 patients, an allele frequency of 1.3%. None of the other variants were identified.

The frequency of the CS/CSL associated variants was established for the patients referred to the WMRGL for diagnostic analysis of these genes, in order to more thoroughly determine the frequency in a similar patient group (see Table 5.38). The cumulative incidence of the CS/CSL-associated variants in the local cohort was most similar to the ExAC control population cohort.

Additionally, 29% (n=51) of patients identified with CS/CSL-associated variants had an additional variant identified during diagnostic genetic analysis (see Table 5.39). 9/15 were known pathogenic mutations and 6/15 were initially classed as VUSs.

Table 5.38 Frequency of variants associated with CS/CSL in cohort of patients referred to WMRGL for testing following diagnosis of PHEO/PGL/HNPGL and/or RCC compared to frequency reported in CS/CSL patients and on a variation database (ExAC)

Local PHEO / PGL/ HNPGL patients referred from the commencement of diagnostic testing in 2004 until September 2011. (SDHB – n=870; SDHC – n= 233; SDHD - n=705)

Gene	coding DNA variation	Protein variation	Local PHEO / PGL/ HNPGL MAF	CS/CSL 2008 Paper (246) MAF (n=74)	CS/CSL 2012 Paper (247) PTEN Negative MAF (n=608)	CS/CSL 2012 Paper (247) PTEN Positive MAF (n=444)	ExAC ALL	ExAC NFE	ExAC Homozygotes observed
SDHB	c.8C>G	p.(Ala3Gly)	0.057%	1.351%	0.164%	0.225%	0.436%	0.019%	Yes (AFR)
SDHB	c.9C>G	p.(Arg27Gly)	0.000%	0.000%	0.164%	0.000%	0.004%	0.008%	No
SDHB	c.158G>A	p.(Gly53Glu)	0.000%	0.000%	0.000%	0.225%	0.084%	0.046%	Yes (NFE)
SDHB	c.170A>G	p.(His57Arg)	0.056%	0.000%	0.164%	0.000%	0.072%	0.001%	No
SDHB	c.359A>G	p.(Asn120Ser)	0.000%	0.000%	0.164%	0.000%	0.000%	0.000%	No
SDHB	c.487T>C	p.(Ser163Pro)	1.445%	2.703%	2.138%	2.027%	1.250%	1.520%	Yes (FIN, NFE, AFR)
SDHB	c.643G>A	p.(Ala215Thr)	0.000%	0.000%	0.000%	0.225%	0.000%	0.000%	No
SDHC	c.197C>T	p.(Ala66Val)	0.000%	0.000%	0.164%	0.000%	0.002%	0.000%	No
SDHD	c.34G>A	p.(Gly12Ser)	0.861%	5.405%	2.961%	0.676%	0.730%	1.020%	Yes (NFE)
SDHD	c.149A>G	p.(His50Arg)	0.780%	2.703%	1.974%	2.477%	0.651%	0.928%	Yes (NFE)
SDHD	c.433C>A	p.(His145Asn)	0.000%	1.351%	0.164%	0.000%	0.000%	0.000%	No
		SUM	**3.200%**	**13.514%**	**8.059%**	**5.856%**	**3.230%**	**3.541%**	

Table 5.39 Summary of patients identified with a CS/CSL and another variant or mutation during this thesis and by the WMRGL since 2004
All patients identified with a CS/CSL-associated variant and another variant are tabulated. LOH analysis is also recorded where appropriate

Number	CS/CSL associated variant identified	LOH analysis performed?	Additional variant identified	Class	LOH performed?
1	*SDHB* c.170A>G p.(His57Arg)	No	*SDHB* c.380T>G p.(Ile127Ser)	Pathogenic (V)	No
2	*SDHB* c.487T>C p.(Ser163Pro)	No	*SDHB* Exon 1 and 2 duplication	VUS (III)	No
3	*SDHB* c.487T>C p.(Ser163Pro)	No	*SDHD* c.191_192delTC p.(Leu64Profs*4)	Pathogenic (V)	No
4	*SDHB* c.487T>C p.(Ser163Pro)	No	*SDHB* c.379dupA p.(Ile127Asnfs*28)	Pathogenic (V)	No
5	*SDHB* c.487T>C p.(Ser163Pro)	Yes - No LOH	SDHC c.380A>G p.(His127Arg)	VUS (III)	Yes - Failed
6	*SDHB* c.487T>C p.(Ser163Pro)	Yes - Variant allele is lost (partial)	*SDHB* c.220G>A p.(Asp74Asn)	VUS (III)	Yes - WT allele is lost (partial)
7	*SDHD* c.34G>A p.(Gly12Ser)	No	*SDHB* c.287-1G>C	Pathogenic (V)	No
8	*SDHD* c.34G>A p.(Gly12Ser)	No	*SDHB* c.137G>A p.(Arg46Gln)	Pathogenic (V)	No
9	*SDHD* c.34G>A p.(Gly12Ser)	No	*SDHB* c.79C>T p.(Arg27*)	Pathogenic (V)	No
10	*SDHD* c.34G>A p.(Gly12Ser)	Yes - No LOH	*SDHB* c.298T>C p.(Ser100Pro)	VUS (III)	Yes - WT allele is lost (total)
11	*SDHD* c.34G>A p.(Gly12Ser)	Yes - No LOH	*SDHB* c.529C>G p.(Arg177Gly)	VUS (III)	Yes - WT allele is lost (total)
12	*SDHD* c.149A>G p.(His50Arg)	Yes - No LOH	*SDHB* c.587G>A p.(Cys196Tyr)	Pathogenic (V)	Yes - WT allele is lost (partial)
13	*SDHD* c.149A>G p.(His50Arg)	No	*RET* c.1826G>A p.(Cys609Tyr)	Pathogenic (V)	No
14	*SDHD* c.149A>G p.(His50Arg)	Yes - No LOH	*SDHB* c.590C>G p.(Pro197Arg)	Pathogenic (V)	Yes - WT allele is lost (total)
15	*SDHD* c.149A>G p.(His50Arg)	No	*SDHB* c.689G>A p.(Arg230His)	VUS (III)	No

5.6.3 Tumour analysis of higher frequency variants

Tumour analysis was performed in all the cases where material could be obtained. For *SDHB* c.487T>C p.(Ser163Pro), 7 FFPE tumour samples were identified. 5/7 patients only had *SDHB* c.487T>C p.(Ser163Pro) and none of the samples showed evidence of loss of heterozygosity (LOH). IHC was performed for three of the samples and all showed preserved expression of SDHB. One patient also had a variant in *SDHC* and, whilst no LOH was observed for *SDHB* c.487T>C p.(Ser163Pro), the SDHC analysis failed to produce a result despite repeated attempts (see Table 5.39). The final patient identified also had a VUS in the same gene, *SDHB* c.220G>A p.(Asp74Asn). That sample showed partial loss of heterozygosity with retention of variant allele for *SDHB* c.220G>A p.(Asp74Asn) and reciprocal loss of the variant allele for *SDHB* c.487T>C p.(Ser163Pro) (see Figure 5.19), showing that they were on opposite alleles and that *SDHB* c.487T>C p.(Ser163Pro) was unlikely to be driving tumourigenesis .

For *SDHD* c.34G>A p.(Gly12Ser), 3 FFPE tumour samples were identified. 1/3 patients only had *SDHD* c.34G>A p.(Gly12Ser). Two also had a Class III VUS in *SDHB*, *SDHB* c.298T>C p.(Ser100Pro) and *SDHB* c.529C>G p.(Arg177Gly). None of the samples showed LOH for *SDHD* c.34G>A p.(Gly12Ser). Both additional VUSs showed LOH, with loss of the normal allele.

For *SDHD* c.149A>G p.(His50Arg), two FFPE tumours were identified. Both patients also had a pathogenic mutation *SDHB* c.587G>A p.(Cys196Tyr) and *SDHB* c.590C>G p.(Pro197Arg). Both samples showed LOH for the pathogenic mutation but not for *SDHD* c.149A>G p.(His50Arg).

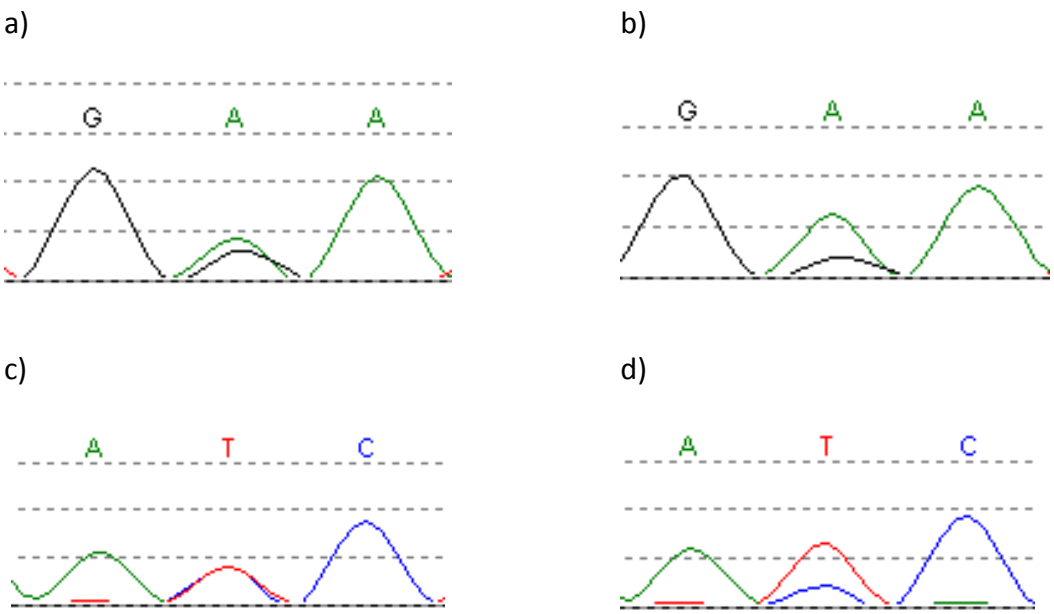


Figure 5.19 LOH analysis for patient with two variants in *SDHB* c.220G>A p.(Asp74Asn) and c.487T>C p.(Ser163Pro)

a) *SDHB* c.220G>A p.(Asp74Asn) in lymphocyte DNA, b) *SDHB* c.220G>A p.(Asp74Asn) in FFPE tumour, c) *SDHB* c.487T>C p.(Ser163Pro) in lymphocyte DNA and d) *SDHB* c.487T>C p.(Ser163Pro) FFPE tumour. The FFPE tumour sequences show that the variants are on opposite alleles as the variant allele is retained for *SDHB* c.220G>A p.(Asp74Asn) and lost for *SDHB* c.487T>C p.(Ser163Pro), indicating that *SDHB* c.220G>A p.(Asp74Asn) is more likely to be involved in tumourigenesis.

5.6.4 Conclusions regarding pathogenicity of these variants

The decision regarding the pathogenicity of the variants discussed in section 5.6 is highly complex. The evidence from the functional analyses performed by Ni *et. al.* together with the increased frequency of certain tumours in the CS/CSL patients with these variants is apparently compelling. However the identification of these variants in patients with *PTEN* mutations suggests that they could be modifiers rather than initiators of disease and that the *PTEN* mutation negative cases may have a yet undiscovered germline initiator mutation. The increased frequency of some of the variants (those at already relatively high frequency) in cohorts with MEN2A and MTC supports the theory that these variants could increase disease penetrance for certain tumour types.

Whilst it was difficult to know how best to report these variants in a patient with symptoms of CS, it had become impossible to know how to report them in patients with PHEO/PGL/HNPGL. This is of particular importance for those variants which are seen at greater than 1% allele frequency, as inherently greater than 1% of patients tested will be found to have them.

The cumulative evidence provided above supports the view that the higher frequency variants are not pathogenic with regard to familial paraganglioma (PHEO/PGL/HNPGL/RCC) tumorigenesis, in the context of causing high risk of cancer:

- The variants are not enriched for in the familial paraganglioma cohort tested at the WMRGL, suggesting they are unlikely to be drivers or indeed modifiers of disease;

- None of the variants were on the retained allele in any patient when LOH analysis was performed. There was either no LOH observed or the variant was on the lost allele;
- IHC analysis of *SDHB* c.487T>C p.(Ser163Pro) found normal staining;
- On nine occasions, a CS/CSL-associated variant was identified with a pathogenic mutation in the same gene, one of the other genes in the SDH complex or another gene which explains the phenotype;
- A CS/CSL-associated variant was identified with a VUS in the same gene or another gene in the SDH complex on six occasions and in all cases where LOH analysis was successfully performed (n=3) LOH was observed with the non-CS/CSL associated variant allele being retained in the tumour, increasing its likelihood of pathogenicity.

Additionally, whilst full phenotype data is not available for the cohort, only 1/51 patients in which one of these variants was identified was reported to have any Cowden syndrome like features. Ni *et. al.* state that the penetrance is low for these variants (246) but likely penetrance of only 2% doesn't fulfil Rahman's criteria for assigning tumour suppressor gene status to a gene (with regard to the CS/CSL phenotype) (2).

From the evidence available it seems appropriate that *SDHB* c.487T>C p.(Ser163Pro), *SDHD* c.34G>A p.(Gly12Ser) and *SDHD* c.149A>G p.(His50Arg) now be reported back to clinicians as not associated with a familial paraganglioma phenotype and that their relevance regarding CS/CSL be reported as remaining uncertain.

5.7 Conclusion

This chapter exemplifies the vast workload that is associated with the classification of variants identified in the course of genetic analysis. Variation databases play a huge role in the classification of variants and even within the comparatively short time between the completion of the gene screening portion of this work and the variant analysis section, the release of the ExAC database allowed reclassification of a number of variants. A set of RNA assays for use with the VHL gene were established which allowed a variant to be reclassified into a clinically actionable category.

In addition, the chapter has demonstrated that functional assays are difficult and labour-intensive to set up and, as shown by the CS/CSL data, even when they show a variant to be associated with altered function, the data must be assessed in context, e.g. the phenotype of all individuals with the variant.

In total, 9/52 (17.3%) variants were reclassified.

Chapter 6 Discussion

6.1 Summary of thesis

Genetics plays a fundamental role in healthcare. One particular aspect of that role is the use of genetics in the context of familial cancer syndromes, specifically in the diagnosis and management of individuals (and their families) who have inherited a mutation in a cancer predisposition gene. Genetic techniques are rapidly advancing and the application of these advancements to the healthcare setting allows more individuals to be diagnosed as having a genetic illness, and hence for better care to be provided for both them and their relatives.

Chapter 3 describes the optimisation, validation and prospective use of Next Generation Sequencing (NGS) for genetic diagnosis of patients with phaeochromocytoma/paraganglioma and renal cell carcinoma. These two overlapping assays allowed increased detection of germline mutations at a lower cost per gene and reduced processing time compared to previous methods of analysis.

As is well known, routine diagnostic genetic analysis can miss pathogenic mutations since it is almost always limited to identifying heterozygous mutations in the coding region of the gene. Chapter 4 describes the optimisation, validation and prospective use of Next Generation Sequencing (NGS) to look for mutations in the full VHL gene region and for mutations occurring at a mosaic level in the tissue that the patient DNA was extracted from. Eighteen percent of clinically diagnosed VHL patients screened using this assay had a clinically actionable mosaic mutation identified.

Unfortunately, the analysis of more genes and fuller analysis of gene regions inherently increases the identification of variants of unknown significance (VUSs). The classification of these VUSs as either benign or pathogenic is very important for the appropriate clinical

management of the individuals that carry them. Chapter 5 describes the attempt to classify the VUSs identified during the course of Chapters 3 and 4 of this thesis, with the addition of some other VUSs identified in the genes studied during previous routine Sanger-based diagnostic analysis by the WMRGL. The chapter demonstrates the many difficulties in variant classification, however, it was possible to reclassify 17.3% of VUSs identified in Chapters 3 and 4 and additional evidence was gathered about those which could not be reclassified.

6.2 The current landscape of genetic medicine

Even in the short time since the commencement of this project there have been significant advances in the field of diagnostic genetics. All British NHS diagnostic genetics laboratories have implemented NGS into routine practice, mainly in the form of gene panel tests such as those described in Chapter 3. Furthermore, exome sequencing is now available from a select set of British diagnostic genetics laboratories with the findings used in the clinic.

There are two ways in which exome sequencing has been implemented. In the first the exome is sequenced but a list of genes that are of relevance to the patient's phenotype is curated and data analysis is restricted to the genes identified, creating a virtual gene panel. This method allows numerous different virtual gene panel tests (more than one gene analysed simultaneously) to be performed using a single technical workflow (254;255). Currently the vast majority of NGS gene panel tests are performed using a specific target enrichment designed in a bespoke manner to a single panel of genes related to a single phenotype or set of overlapping phenotypes, such as the panels in Chapter 3. The new model can be implemented either with whole exome enrichment or with smaller

enrichment limited to those genes in the exome already known to be associated with genetic disease (sometimes referred to as the 'Mendeliome'). Whilst this model has little scope for novel disease gene identification, it is a cost effective and streamlined method of gene panel analysis. Additional benefits are that the genes in a panel can be easily updated and that the exome data for a patient can be stored and analysed for additional genes or using a new gene list for a different phenotype, should the need arise (254). One limitation of this method is that the exome enrichment may not fully cover the coding region of all of the genes of interest, thus decreasing the clinical sensitivity of the test compared to a specifically designed panel. In a study using this model, with the Agilent Sure Select All Human Exon Kit for target enrichment, it was estimated that only ~90% of genes were reliably covered (255). Another limitation is that a laboratory routinely performing a specific panel test becomes expert in the interpretation of its results and, given that interpretation of VUSs is central to diagnostic genetics, a movement away from centres with expert knowledge could hinder VUS interpretation.

The second way in which exome sequencing has been implemented is following the model used in research for novel gene identification first demonstrated by Ng *et. al.* (256). In this approach, all variants identified within the exome enrichment used are reported and then prioritised using a locally validated protocol. An example of a flowchart depicting one such protocol published by Baylor College of Medicine, USA is shown in Figure 6.1 (257). A number of publications have demonstrated how this approach can work in the diagnostic setting (257-260). For example, the Canadian Finding Of Rare Disease Genes study showed how exome sequencing could be used for both novel gene identification and genetic diagnosis in patients who have mutations in previously characterised genes, but where the

causative mutation(s) had not been identified by previous analysis (260), with the latter being the more likely scenario in a diagnostic setting in the short term. That study identified a mutation in a known gene in 105/362 families analysed (260). The lack of previous identification of the pathogenic change(s) was most commonly the result of either a high level of genetic heterogeneity related to the patient's condition or an atypical presentation (260). One of the disadvantages of whole exome sequencing is the increased burden of variant interpretation. It has been established that each sequenced exome has 20, 000 to 50, 000 variants identified and that following variant prioritisation methods, such as that in Figure 6.1, there are still between 150 and 500 private non-synonymous or splice-site variants requiring further analysis and interpretation in the context of the phenotype (261). It should also be noted that the pathogenic mutation could have been filtered out during the prioritisation process (261).

The majority of the published studies in this area are for patients with childhood-onset disorders and do not include cancer patients. In one study where cancer patients were included and which used the first model (virtual panel), a diagnostic yield of only 3% was achieved when the authors looked at a group of patients with microsatellite stable colorectal cancer, in comparison to a diagnostic yield of 52% in their blindness cohort (255). However, it is hoped that the use of clinical whole exome analysis could identify some currently unknown cancer predisposition genes (CPGs) and additional tumour types associated with known CPGs. In addition, CPG-specific enrichments are available, such as the Illumina TruSight Cancer enrichment. This means a patient with inherited cancer can be analysed for the vast majority of known CPGs, thus identifying patients with atypical presentation of disease.

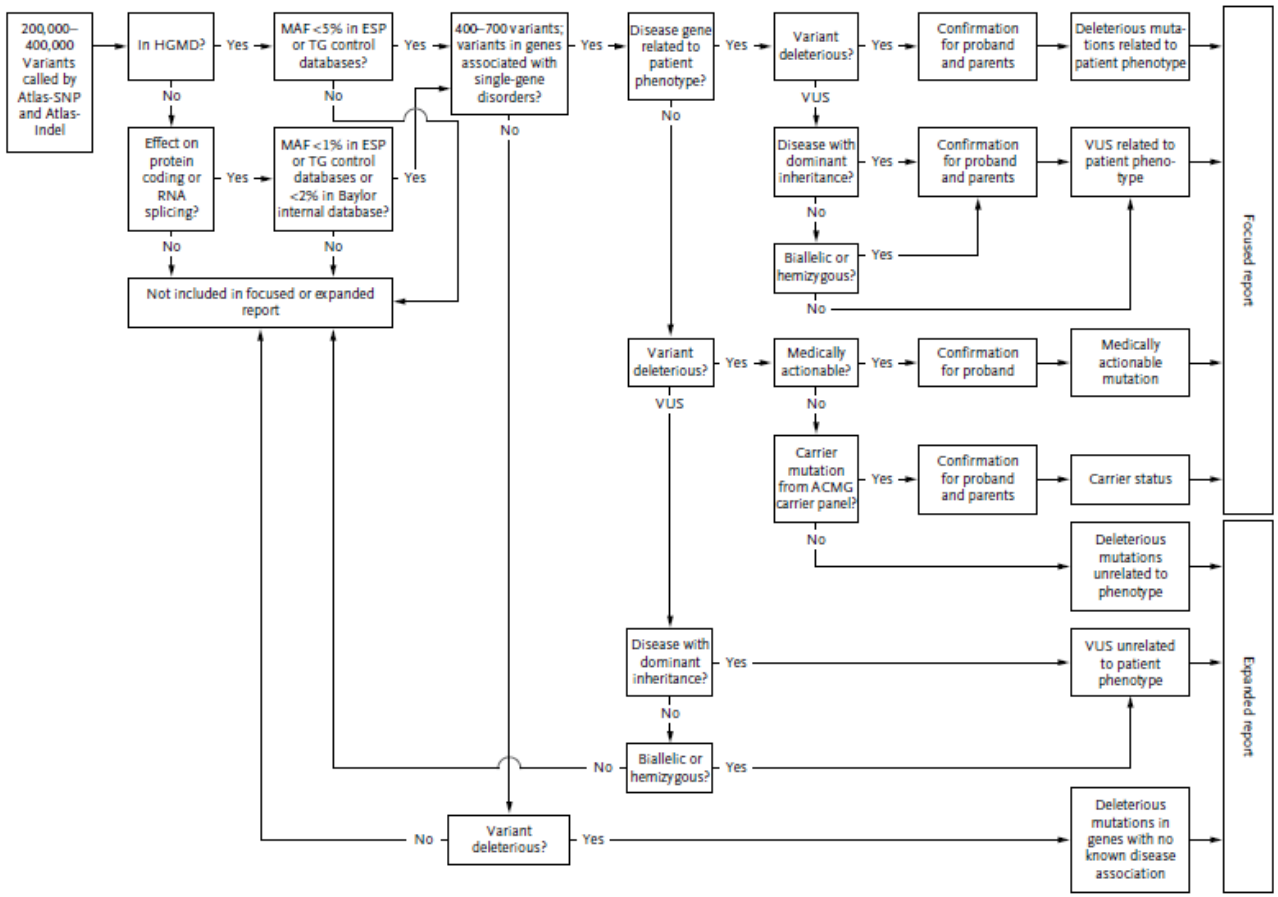


Figure 6.1 Example variant prioritisation and interpretation protocol published by Yang *et. al.*
ESP = ESP5400 data from the National Heart, Lung, and Blood Institute GO Exome Sequencing Project, HGMD = Human Gene Mutation Database, MAF = minor allele frequency, TG = the 1000 Genomes Project and VUS = variants of unknown clinical significance. (Adapted from (257)).

The use of exome sequencing brings with it the issue of incidental findings (IFs). Green *et. al.* defined IFs as 'results that are not related to the indication for ordering the sequencing but that may nonetheless be of medical value or utility to the ordering physician and the patient' (262). There has been much discussion in the literature regarding the ethics and the utility of reporting such findings. In 2013, the American College of Medical Genetics and Genomics (ACMG) released their recommendations for dealing with IFs (262). In brief, they recommended that mutations in a defined list of 57 genes should not only be reported but the data be sought if it was available from the analyses performed (262), e.g. exome sequencing. The list mainly comprised genes in which pathogenic mutations could be life-limiting, such as genes related to cardiovascular diseases and cancer predisposition (262) and included the majority of the genes studied during this research. The recommendations included that those being tested would be obliged to have any incidental findings reported to them (262).

There was a flurry of responses to the recommendations, with critics suggesting the ACMG guidelines ignored the patient's right for autonomy (263). Various other internationally recognised bodies released their versions but none proposed such definite ideas as the American group (264-267). In fact, a 2014 report from the European Society for Human Genetics recounted that during their annual conference debate there was no audience consensus regarding whether incidental findings should be reported back to patients or not (268), let alone actively sought as suggested by the ACMG (262). The debate over incidental findings is further complicated by VUSs and how and when they should be reported to patients if identified in genes such as those highlighted by the ACMG.

Another difficulty with the rapid technological advancement in genetic medicine is deciding which test(s) will be most useful for which patients, whilst achieving a diagnosis at the lowest cost possible. There is still much value in traditional genetic analysis, such as single gene analysis and chromosome analysis. A retrospective study of the genetic diagnosis of 500 patients at Duke University Medical Center, USA found that 46% of patients were diagnosed using these traditional methods and that it therefore was not necessary or cost effective to request complex tests in these cases (269). The advantages and disadvantages of each type of test must be recognised, see Table 6.1, and related back to the patient and their phenotype. Xue *et. al.* have proposed a general algorithm to aid in the clinician's decision making process, see Figure 6.2 (270).

Table 6.1 Factors to consider when selecting a sequencing based diagnostic test.
CDS = coding sequence, VUS = variant of unknown significance, IF = incidental finding. (Based on (270)).

Test	Coverage of CDS for gene(s) of interest	Chance of VUS identification	Chance of IF identification	Cost	Additional test(s) required
Single gene test	Comprehensive	Minimal	Minimal	Low	None
Gene panel test	Moderate to comprehensive	Moderate	Low	Moderate to high	Sanger sequencing required to confirm variants. May also be used to fill coverage gaps
Exome test	Variable	High	High	High	Sanger sequencing required to confirm variants

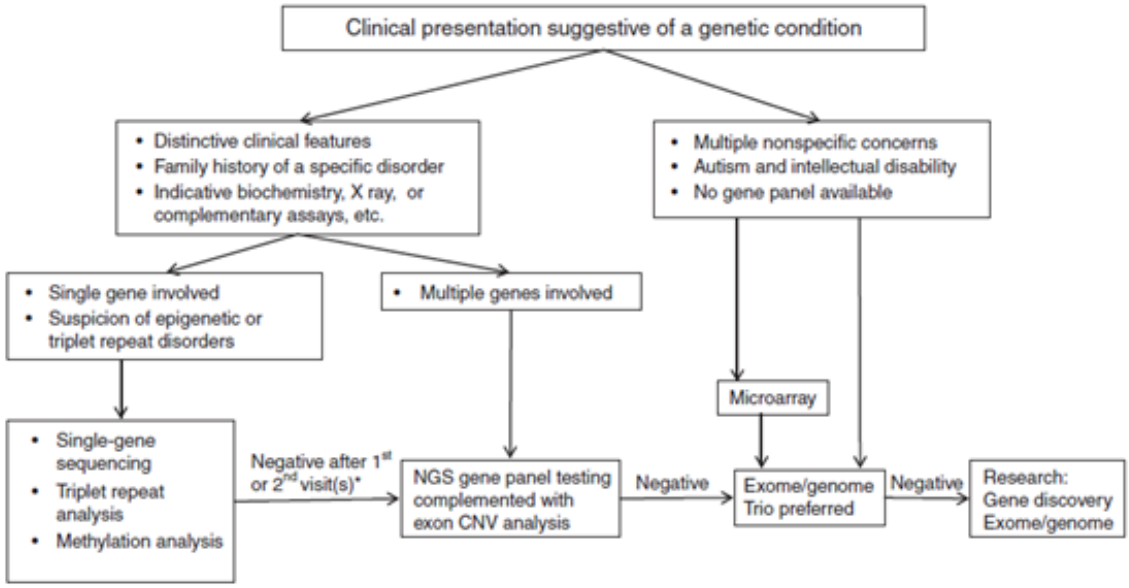


Figure 6.2 Molecular genetic testing algorithm
(Adapted from (270)).

In terms of the cancer predisposition diseases examined in this research, the current optimum testing strategy is highly dependent on the specific patient. For example a patient with a clinical diagnosis of VHL disease Type 1 would still currently be best served by either single gene analysis or a comprehensive NGS panel depending upon cost, followed by a test allowing mosaic level examination if no mutation was identified. In contrast, a patient with inherited paraganglioma would be best served by a comprehensive NGS panel, followed by exome analysis for novel gene identification, if required.

6.3 The future of genetic medicine

Whole genome sequencing is considered likely to become the test of choice in coming years (271). The British government has recognised this and has commissioned the 100, 000 genomes project to be run by a Department of Health owned company called Genomics

England (272). This project aims to perform whole genome sequencing on 100, 000 NHS patients by 2017, focussing on patients with rare diseases and cancer (272). For patients with rare diseases they plan to sequence trios (patient and both parents), whilst for cancer patients they plan to sequence germline and tumour DNA samples (272). Genomics England recognises that there are three main challenges involved with a project of this magnitude, sequencing, data and security (272). The sequencing challenge concerns ensuring the quality of the amount of sequencing to be performed; initially all findings will be confirmed in a diagnostic laboratory (272). Data refers to both bioinformatic processing and the clinical interpretation of the variants identified (272). Security refers to the appropriate storage of the data obtained (272). In particular recognition of the data interpretation challenges, a number of Genomics England Clinical Interpretation Partnership (GeCIP) domains have been created (272). These GeCIPs are described as having three main aims:

'1. To optimise clinical data and sample collection, clinical reporting and data validation and interpretation.

2. To improve understanding of the implications of genomic findings and improve the accuracy and reliability of information fed back to patients. To add to knowledge of the genetic basis of disease.

3. To provide a sustainable thriving training environment.' (272).

Therefore, Genomics England recognises the vast problems they face with variant pathogenicity determination, as exemplified on a smaller scale by Chapter 5 of this research.

Interestingly, the 100, 000 genomes project has decided on an 'opt-in' policy with regard to incidental findings (272). In the context of the project they are referred to as secondary findings and defined as 'an additional looked-for health-related finding that is not pertinent to the main condition' (272). A patient has to opt-in to having secondary findings actively sought and reported back from a list of defined genes/conditions known to influence healthcare, similar to those included on the ACMG list (262;272). VUS in these genes will not be reported, however reanalysis of genes which are later identified to meet these criteria is included (272), as, it must assumed, variants will be variant with upgraded pathogenicity status, at least within the context of the study.

6.4 Final conclusions

This thesis describes the identification and assessment of variants in a subset of genes involved in familial cancer syndromes. The development of three novel variant identification assays has allowed genetic diagnosis in patients who would have otherwise remained uncertain of their family's genetic status. This thesis has also demonstrated the difficulty in successful variant pathogenicity determination, both in the context of highly penetrant genetic diseases and genetic modifiers. The meteoric rise in both the importance of genetics in healthcare and the vast improvements in the technology available to perform genetic analyses, mean that variant pathogenicity determination will remain at the centre of diagnostic genetics.

Chapter 7 Appendices

7.1 Chapter 3 Appendices

Table 7.1 Primer sequences

Primers were tagged with the Fluidigm CS tags for NGS and the M13 tags for Sanger confirmations and gap-fills. CS forward tag: ACACTGACGACATGGTTCTACA, CS reverse tag: TACGGTAGCAGAGACTTGGTCT, M13 forward tag: TGTAAAACGACGGCCAGT, M13 reverse tag: CAGGAAACAGCTATGACC.

Gene_Exon	F Primer	R Primer	Chromosome	Product size
FH_1	GGTCATTCACGATTGGTCCT	GAATCTCTCCCGCCAAGTC	1	375
FH_10	TTAAGAGGCCCATATAGCATCAA	TCCATCTTAGACCTAGCACATCC	1	422
FH_2	CAGCTGATAAGATGCGATTACTTT	TATTCAAAGCGATGGCTCAA	1	419
FH_3	TCTTGGTTCATGAATTCCTGTCT	TGCCAGAGCATATCGTCATC	1	360
FH_4	TCTGTGGTTGGAGCAAGTGA	AAGAACCATAAGAAGCCTTATCCA	1	434
FH_5	ACCAGGATGCTGCTTATTTGA	ATTGGCCATTTGTACCAAGC	1	419
FH_6	CCCTCATCCTTCCCTATACTTTG	CACAAGAATTCAAGACAGGAACA	1	436
FH_7	TGTTTCACTTGCTAATGGTAGAAAA	TGACCAGAGGACCACAGACA	1	428
FH_8	CTACCCATCCCACCTTCCTT	ACCCAACTACCCAATGTGGA	1	397
FH_9	GCCTTCAAATGTTCATGCTTT	GCTGTTCTCAAACACTGATCCA	1	450
FLCN_4	GGGAGGTTTCATGGAGTCAA	CCTGTCCATCCCACACCTAC	17	390
FLCN_5	TTCCGAGCTCAGATTTGCAT	GCTCAGCAAGTCCAACATGA	17	432
FLCN_6	TGGTGTCACTAAGCGAGGAA	GTGCACTGGCTGTAAGCAGA	17	393
FLCN_7	GATTGGACTGGTGGCACTTT	TGGACAAGCCAACCAATGTA	17	388
FLCN_8	GTGAGCGTCAGGTTTGCTTT	GACAGCCCACCTGTGAATTT	17	412
FLCN_9	GGCTCAGCCCATGAAGTATC	GAAGGTGGAGGGTCCAGAG	17	371
FLCN_10	CTGAGCCCTGTCTTTGCTCT	GCACACGCATCCTTCTGA	17	425
FLCN_11	ACAAGCTGGTGTGTGACTGG	TCCACAACCCATGACAGAGA	17	446
FLCN_11a	CACTGTGGGCTGAGAGTCTG	CACCGCTACCTGTGTGAAGA	17	448

Gene_Exon	F Primer	R Primer	Chromosome	Product size
FLCN_12	CATTGATTTGGGCCAGTCTT	CCGCTTCAATCTTATTCAGGA	17	398
FLCN_13	GAGTTTGCTGTCATCGTGGA	TTGGAAACAGCTCCAGGTTT	17	399
FLCN_14	GGATTGTGCTGTGGTGTCTG	CCTGATGGTTTCCCTTCCTT	17	380
MAX_1	GCCGTCTGTTAGAGGGACAA	GAGGGGAAGGGGAAGGAG	14	445
MAX_2	GCCGGCATTTTCTTTCTTATT	GGCATCTGGATCCCCTATCT	14	392
MAX_3	TGCTGACAGCGTGTACTTCC	TGCAGTGATCCTCCAAACAG	14	357
MAX_4	TTCAGCCTTTTCCCACCTTA	CCTCAGGACGGCTCTAACAC	14	416
MAX_5	CAACAAGCAAGCCCTTAGGA	GGAGGATGAGACGATGGAGA	14	348
MET_16	AATGAAGCTCATAAAGGGTTTGA	AGGGCTCTGAGGGATCATTT	7	364
MET_17	AAACCCTCAGGACAAGATGC	AGGGATGGCTGGCTTACAG	7	360
MET_18	GGCTTGAGCCATTAAGACCA	CTGAAAGTCCCACCCACATC	7	413
MET_19	GGATTTCAAATACTGAAGCCACTT	CTGGAATTGGTGGTGTTGAA	7	379
RET_10	TGGAGGCAGAGTCCTTTGTT	TGCTGTTGAGACCTCTGTGG	10	423
RET_11	ATACGCAGCCTGTACCCAGT	CACAGGATGGCCTCTGTCTC	10	419
RET_13	TGACCTGGTATGGTCATGGA	ATGGCAGTGTCACACCAGAG	10	450
RET_14	AAGACCCAAGCTGCCTGAC	GGCTAGAGTGTGGCATGGT	10	383
RET_15	CTGGTCACACCAGGCTGAG	AGAGGCTGAGCGGAGTTCTA	10	386
RET_16	CATCTCAGCAATCCACAGGA	GCTAGCACTGCAGACAGGTG	10	403
SDHA_1	GACCTCAGCGTTCCCTTAAC	GCTCTCTCAGCCCCTTCC	5	450
SDHA_2	CAAGTCAGGGTGGAAACGAG	CCTTACCCCCTAAGCCAAAA	5	362
SDHA_3	TTGGCATAGTGGAACATGTGA	AAGAGGCTCCAGGGAGAGAC	5	374
SDHA_4	TCTGGAATCTGTCGGGTCTC	CCGCCTCAATTGCTTACATT	5	397
SDHA_5	TTTGGGTTTGCAGATTTGTG	CACGCTGCTGTTCTCTGTTG	5	399
SDHA_6	TGGTCCATTTGGATCAAGTTC	CCATCACCCTCACATGGTTA	5	397
SDHA_7	TGTGCACAGCACTGAGAAGAC	CCTGAGAGCTGAGGAAGCTG	5	374
SDHA_8	TGTCAGCCTTGTCAGTGCTT	CGGTGCTGAAACTCACAGAA	5	436

Gene_Exon	F Primer	R Primer	Chromosome	Product size
SDHA_9	GATTTGATGAGAGGGTGACCA	GTCGTCCCAGCCTTCTTTC	5	392
SDHA_10	GGCACCTTGACATTTCACCT	TGGGACTCTGAAGGTTGAGG	5	383
SDHA_11	TGGCAGTGTGTTAGCTCAGG	GACTCCACCACCTCCCAGTA	5	395
SDHA_12	CAGTTCTTGGTATGGCTGCTT	CATGGAGTCACTGGTTGTGG	5	324
SDHA_13	GTCCTGGCCCACAGCTATAA	TTTACTTTTGGGGCAACGAA	5	366
SDHA_14	CCTGGGCTGAAGTGGAAAT	CGAATCTGTCGGGAGTTGAT	5	369
SDHA_15	AAGGAACACAAGAGATGGCTTT	AATGAAGCAAGGGACAAAGG	5	362
SDHAF2_1	CCGTTACCGGGAATATGG	TATCGGGCAGACGAACTCTC	11	379
SDHAF2_2	CCTGGCCAGCAGTGTAATTT	CCTACTCTGGCTCAGGAAGC	11	410
SDHAF2_3	GCAGAAAGAGGGGAATGTTG	GGCAACGAGAGTGAAACTCA	11	433
SDHAF2_4	GTGGTTCTTGGCCAGTGTTT	TGAGGCTCAGAGCTGAACAA	11	439
SDHB_1	CGGCTAGTGGGTCCTCAGT	ACTTGACAGGGATGCTCGAA	1	350
SDHB_2	TCCAGCGTTACATCTGTTGTG	GGCCACATTGGTCTTGAACT	1	447
SDHB_3	CCGAAGGTGACCTGAGAAGA	GCCAGATGCAGTAACCACAG	1	449
SDHB_4	CAGCAAGGAGGATCCAGAAG	CTCTGGCCACACTGTCTCAG	1	448
SDHB_5	CCTGGCATAGAGTGGACGAG	AGGGCTCTAGCTCCTTGGTC	1	445
SDHB_6	TTAAGAACCCTGGGCAGATG	ATGGCAATGAAGGAAACCAG	1	434
SDHB_7	CCCAGAGCTTTGAGTTGAGC	TCCCACAATGTCCCTGTGTA	1	444
SDHB_8	CACCTTGCTTGGACACTGAA	GACTCCTGGCACCTTCACAT	1	397
SDHC_1	CCGAAATCCTGGCAAATTC	CTGCCCAGGCACAGGATA	1	352
SDHC_2	TTGTCTTGCATTATTTGTGTTTGA	CTGGGTGACAGAGCGAGACT	1	368
SDHC_3	GATTACAGGCCTGAGCAACC	TCTACCCTGAAGGGTTCACCT	1	368
SDHC_4	ACAGGAATGCAAAAGCTGGT	TCCCTCTAAATCAAGTGCTGAGT	1	450
SDHC_5	GGAGACTCGCTTGAACTTGG	AGTCTCCCCACTCCCTTCAC	1	447
SDHC_6	CCTCCTACCATCACCACACA	TCCCAGGCTGGAGATAAGAA	1	429
SDHD_1	ACCTTCCGACAGCTGTGTTT	CCTTCGGGTAAACATCTGGA	11	381

Gene_Exon	F Primer	R Primer	Chromosome	Product size
SDHD_2	ATGCAGATGTTCCCTGGTCT	TGTCTGCCCAAAGGTGTAAA	11	450
SDHD_3	TGCCTGTCAGTTTGGGTTACT	GACATTGAATAGGCAGGCAGA	11	415
SDHD_4	GGAGTGGCAAATGGAGACAT	AAGAAGGCTGTCCACCAATG	11	411
TMEM127_2	TGCTATCCTCCACCGGACT	CTGGTCCCTGGCTATCTCTG	2	421
TMEM127_3	GCTAAGACAGCGTCCTCCAG	AAGACTGGAGCTTGGCACAT	2	431
TMEM127_4	TGCTCACTAATTCCCCTTCTG	GGCTGCCGAGGAAGAGAG	2	381
VHL_1a	AGCGCGTTCCATCCTCTA	GCGATTGCAGAAGATGACCT	3	370
VHL_1b	GGCAGGCGTCGAAGAGTA	GCTTCAGACCGTGCTATCGT	3	345
VHL_2	GGACGGTCTTGATCTCCTGA	TTGAGACACCATAACACCTTTAACA	3	451
VHL_3	GCAAAGCCTCTTGTTCGTTC	AGGAAGGAACCAGTCCTGTATCT	3	338

7.2 Chapter 4 Appendices

Table 7.2 Class I - Benign variants excluded from further analysis during NGS screening of the VHL gene region

These variants have been classed as benign on the basis of minimum allele frequency (MAF). MAF = Minimum allele frequency; 1000 genomes ALL_MAF = MAF of all populations in 1000 genomes phase 1 study; AFR_MAF = MAF of all African individuals in 1000 genomes phase 1 study; AMR_MAF = MAF of all American individuals in 1000 genomes phase 1 study; ASN_MAF = MAF of all East Asian individuals 1000 genomes phase 1 study; EUR_MAF = MAF individuals of all European individuals 1000 genomes phase 1 study. * Seen repeatedly in West Midlands Regional Genetics Laboratory *in trans* with pathogenic mutations.

Number	Chromosome position - start	dbSNP reference	Study MAF	dbSNP MAF	1000 genomes ALL_MAF	1000 genomes AFR_MAF	1000 genomes AMR_MAF	1000 genomes ASN_MAF	1000 genomes EUR_MAF
1	10179646	rs3864019	0.0210	0.0268	G:0.0257	G:0.03	G:0.07	G:0.0017	G:0.02
2	10179778	rs117742184	0.0035	0.0216	T:0.0073	T:0.0020	T:0.0028	T:0.02	-
3	10179862	rs9879931	0.1538	0.222	T:0.2222	T:0.56	T:0.21	T:0.0035	T:0.18
4	10179912	rs60650700	0.0070	0.0312	T:0.0248	T:0.09	T:0.02	-	T:0.0013
5	10179924	rs140812048	0.0490	0.0482	T:0.0482	T:0.04	T:0.03	T:0.07	T:0.04
6	10180041	rs9875667	0.1818	0.2801	C:0.2801	C:0.70	C:0.23	C:0.10	C:0.17
7	10180062	rs14465863	0.0035	0.0078	T:0.0078	T:0.03	T:0.01	-	-
8	10180088	rs191444064	0.0035	0.0078	A:0.0096	A:0.03	A:0.01	-	A:0.01
9	10180557	rs776517	0.1434	0.1221	G:0.1221	A:0.94	A:0.81	A:0.87	A:0.88
10	10180659	rs265318	0.1364	0.1203	C:0.1203	A:0.91	A:0.81	A:0.87	A:0.90
11	10181240	rs713024	0.1294	0.1198	A:0.1198	G:0.94	G:0.81	G:0.87	G:0.88
12	10181245	rs151107472	0.0035	0.0038	G:0.0073	-	-	G:0.03	G:0.0013
13	10181812	rs111860145	0.0035	0.0284	A:0.0220	A:0.08	A:0.02	-	-
14	10181819	rs137893497	0.0035	0.0212	A:0.0073	A:0.0020	A:0.0028	A:0.02	-
15	10182040	rs199563741	0.0035	0.0319	-:0.0294	-:0.11	-:0.02	-	-:0.0026
16	10182150	rs73024525	0.0105	0.003	C:0.0046	-	-	-	C:0.01

Number	Chromosome position - start	dbSNP reference	Study MAF	dbSNP MAF	1000 genomes ALL_MAF	1000 genomes AFR_MAF	1000 genomes AMR_MAF	1000 genomes ASN_MAF	1000 genomes EUR_MAF
17	10182562	rs779803	0.3007	0.3898	G:0.3898	A:0.29	A:0.59	A:0.77	A:0.70
18	10183002	rs10433558	0.2483	0.3214	C:0.3214	C:0.05	C:0.30	C:0.64	C:0.27
19	10183274	rs779804	0.3636	0.135	C:0.1350	T:0.90	T:0.80	T:0.87	T:0.86
20	10183337	rs779805	0.3077	0.4164	G:0.4164	A:0.18	A:0.57	A:0.77	A:0.70
21	10183605	rs35460768 *	0.0035	0.0004	Not available	Not available	Not available	Not available	Not available
22	10183876	rs61758376	0.0070	0.0214	C:0.0073	C:0020	C:0.0028	C:0.02	-
23	10184255	rs779806	0.1608	0.1506	G:0.1506	A:0.86	A:0.78	A:0.87	A:0.86
24	10184519	rs73024533	0.0070	0.0026	C:0.0046	C:0.0020	C:0.0028	-	C:0.01
25	10184865	rs11920834	0.0035	0.0331	G:0.0289	G:0.11	G:0.02	-	-
26	10184956	rs779807	0.2972	0.4366	A:0.4366	T:0.17	T:0.57	T:0.74	T:0.68
27	10184963	rs191112313	0.0035	0.0212	T:0.0073	T:0.0020	T:0.0028	T:0.02	-
28	10185021	rs779808	0.2972	0.4164	C:0.4164	T:0.18	T:0.57	T:0.77	T:0.70
29	10185128	rs73024535	0.0105	0.003	G:0.0046	-	-	-	G:0.01
30	10185262	rs779809	0.1294	0.1217	A:0.1217	G:0.93	G:0.81	G:0.87	G:0.88
31	10185354	rs524411	0.0350	0.0638	G:0.0638	G:0.01	G:0.08	G:0.12	G:0.05
32	10185803	rs146851054	0.0035	0.0212	C:0.0073	C:0.0020	C:0.0028	C:0.02	-
33	10186216	rs34232505	0.1434	0.1993	A:0.1993	A:0.41	A:0.20	A:0.08	A:0.16
34	10186268	rs74749021	0.0035	0.007	T:0.0078	T:0.03	T:0.01	-	-
35	10186447	rs1678593	0.3042	0.4192	A:0.4192	T:0.18	T:0.57	T:0.77	T:0.70
36	10186514	rs779811	0.3042	0.4669	T:0.4669	A:0.16	A:0.54	A:0.70	A:0.65
37	10187014	rs779812	0.1364	0.1212	A:0.1212	G:0.93	G:0.81	G:0.87	G:0.88
38	10187112	rs779813	0.1364	0.1212	A:0.1212	G:0.93	G:0.81	G:0.87	G:0.88
39	10187171	rs374645	0.2063	0.0693	T:0.0693	C:0.95	C:0.89	C:0.97	C:0.91

Number	Chromosome position - start	dbSNP reference	Study MAF	dbSNP MAF	1000 genomes ALL_MAF	1000 genomes AFR_MAF	1000 genomes AMR_MAF	1000 genomes ASN_MAF	1000 genomes EUR_MAF
40	10187175	rs113225030	0.0035	0.0034	T:0.0032	T:0.01	-	-	-
41	10187369	rs181381146	0.0035	0.0198	C:0.0073	C:0.0020	C:0.0028	C:0.02	-
42	10187377	rs1642729	0.1993	0.1717	A:0.1717	G:0.87	G:0.76	G:0.83	G:0.83
43	10187634	rs2543458	0.1364	0.1217	G:0.1217	T:0.93	T:0.81	T:0.87	T:0.88
44	10187690	rs76843316	0.1608	0.3127	-:0.3127	C:0.57	C:0.68	C:0.74	C:0.73
45	10187884	rs9819196	0.0105	0.0763	A:0.0666	A:0.25	A:0.03	-	A:0.02
46	10187949	rs371330108	0.0070	0.0194	Not available	Not available	Not available	Not available	Not available
47	10187952	rs149056499	0.0105	0.003	T:0.0051	-	-	T:0.0017	T:0.01
48	10187982	rs143057620	0.0490	0.0184	T:0.0184	T:0.01	T:0.02	-	T:0.04
49	10188011	rs35286098	0.1608	0.2544	C:0.2544	-:0.45	-:0.74	-:0.85	-:0.86
50	10188363	rs34661876	0.0175	0.0048	G:0.0083	-	G:0.01	-	G:0.02
51	10188428	rs1678607	0.1364	0.208	T:0.2080	G:0.88	G:0.75	G:0.75	G:0.79
52	10188494	rs140172229	0.0035	0.0765	-:0.0666	-:0.25	-:0.03	-	-:0.02
53	10188576	rs143659897	0.0035	0.0022	T:0.0023	T:0.01	-	-	-
54	10188798	rs1678606	0.1364	0.1244	T:0.1244	C:0.93	C:0.81	C:0.87	C:0.87
55	10188835	rs62238314	0.2517	0.32	G:0.3200	G:0.05	G:0.29	G:0.64	G:0.27
56	10188884	rs112248386	0.1434	0.1212	T:0.1212	T:0.09	T:0.16	T:0.08	T:0.16
57	10189493	rs9829290	0.0035	0.1274	A:0.1010	A:0.36	A:0.04	A:0.02	A:0.02
58	10189929	rs1681660	0.1434	0.1217	T:0.1217	C:0.93	C:0.81	C:0.87	C:0.88
59	10189935	rs147654121	0.0035	0.0088	T:0.0055	T:0.01	-	T:0.02	-
60	10190100	rs189742164	0.0035	0.0767	A:0.0579	A:0.19	A:0.03	-	A:0.03
61	10190293	rs112782301	0.0874	0.124	T:0.1240	T:0.10	T:0.16	T:0.08	T:0.16
62	10190458	rs189039866	0.0070	0.0018	C:0.0018	-	-	-	C:0.01

Number	Chromosome position - start	dbSNP reference	Study MAF	dbSNP MAF	1000 genomes ALL_MAF	1000 genomes AFR_MAF	1000 genomes AMR_MAF	1000 genomes ASN_MAF	1000 genomes EUR_MAF
63	10190458	rs189039866	0.0035	0.0018	C:0.0018	-	-	-	C:0.01
64	10190467	rs1642743	0.3601	0.399	C:0.3990	T:0.30	T:0.57	T:0.78	T:0.67
65	10190577	rs2600006	0.4755	0.3719	A:0.3719	A:0.36	A:0.41	A:0.22	A:0.48
66	10190661	rs2600005	0.1573	0.2456	A:0.2456	G:0.49	G:0.74	G:0.85	G:0.86
67	10190753	rs116127341	0.0175	0.0275	A:0.0275	A:0.0020	A:0.04	-	A:0.06
68	10191219	rs1703153	0.3182	0.4077	C:0.4077	G:0.22	G:0.58	G:0.77	G:0.70
69	10191377	rs116128787	0.0035	0.138	A:0.0115	A:0.04	A:0.01	-	-
70	10191943	rs1642742	0.3182	0.4073	G:0.4073	A:0.22	A:0.58	A:0.77	A:0.70
71	10192672	rs1681669	0.3182	0.3866	G:0.3866	A:0.31	A:0.58	A:0.77	A:0.70
72	10192709	rs1681668	0.3182	0.4054	C:0.4054	T:0.22	T:0.59	T:0.77	T:0.71
73	10192762	rs1642741	0.1503	0.1244	T:0.1244	C:0.92	C:0.81	C:0.87	C:0.88
74	10193509	rs458106	0.0140	0.0161	A:0.0161	G:0.99	G:0.96	-	G:0.98
75	10193622	rs149248243	0.0035	0.12	Not available	Not available	Not available	Not available	Not available
76	10193683	rs1136249	0.3217	0.4169	T:0.4169	G:0.18	G:0.57	G:0.77	G:0.71
77	10193789	rs140614750	0.0175	0.1468	T:0.1129	T:0.41	T:0.04	T:0.03	T:0.02
78	10194243	rs191582744	0.0070	0.0264	A:0.0078	A:0.0020	A:0.0028	A:0.03	-
79	10194249	rs142728549	0.0245	0.0471	A:0.0482	A:0.01	A:0.15	-	A:0.06
80	10194624	rs801913	0.1469	0.1134	G:0.1134	C:0.97	C:0.81	C:0.87	C:0.88
81	10195172	rs17610448	0.0210	0.0128	G:0.0184	G:0.0041	G:0.06	-	G:0.02
82	10195252	rs145137834	0.0070	0.002	T:0.0028	-	-	-	T:0.01
83	11128507	rs11128507	0.0105	0.0771	A:0.0661	A:0.25	A:0.03	-	A:0.02

Table 7.3 Artefacts excluded from further analysis during NGS screening of the VHL gene region

Chromosome location, variant call and possible explanation for false call, with reasoning behind exclusion from further analysis.

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
1	10179447	c.[-4085G>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
2	10179534	c.[-3998A>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
3	10179541	c.-3992_-3991insT;c.-3991_-3990delTT	Local homopolymer	Internal to homopolymer
4	10179552	c.-3980_-3979delTT	Local homopolymer	Internal to homopolymer
5	10179554	c.-3979_-3978insT	Local homopolymer	Internal to homopolymer
6	10179860	c.[-3672T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
7	10179901	c.[-3631T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
8	10181322	c.[-2210A>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
9	10181331	c.[-2201A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
10	10181334	c.-2199_-2198insT;c.-2198_-2197delTT	Local homopolymer	Internal to homopolymer
11	10181334	c.-2199_-2198insT;c.-2198delT	Local homopolymer	Internal to homopolymer

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
12	10181345	c.-2187_-2186delTT	Local homopolymer	Internal to homopolymer
13	10181346	c.-2186delT	Local homopolymer	Internal to homopolymer
14	10181347	c.-2186_-2185insT	Local homopolymer	Internal to homopolymer
15	10182231	c.-1301_-1297delTTTTT	Local homopolymer	Internal to homopolymer
16	10182231	c.-1302_-1301insT;c.-1301_-1297delTTTTT	Local homopolymer	Internal to homopolymer
17	10182244	c.-1288_-1283delTTTTTT	Local homopolymer	Internal to homopolymer
18	10182245	c.-1287_-1283delTTTTT	Local homopolymer	Internal to homopolymer
19	10182249	c.-1283_-1282delTG	Local homopolymer	Adjacent to homopolymer
20	10182250	c.-1283_-1282insT	Local homopolymer	Internal to homopolymer
21	10182250	c.-1282delG	Local homopolymer	Internal to homopolymer
22	10182937	c.-595delA	Local homopolymer	Internal to homopolymer
23	10182937	c.-596_-595insA;c.-595delA	Artefact	Low percentage reads, all samples within 1 or more runs

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
24	10182946	c.-586delA	Local homopolymer	Internal to homopolymer
25	10182947	c.-586_-585insA	Local homopolymer	Internal to homopolymer
26	10183242	c.-291_-290insA;c.-290_-289delAA	Local homopolymer	Internal to homopolymer
27	10183242	c.-291_-290insA;c.-290delA	Local homopolymer	Internal to homopolymer
28	10183242	c.-291_-290insA	Local homopolymer	Internal to homopolymer
29	10183254	c.-278_-277delAA	Local homopolymer	Internal to homopolymer
30	10183255	c.-277delA	Local homopolymer	Internal to homopolymer
31	10183256	c.-277_-276insAA	Local homopolymer	Internal to homopolymer
32	10183256	c.-277_-276insA;c.[-276C>A];[=]	Local homopolymer	Adjacent/Internal to homopolymer
33	10183256	c.-277_-276insAA;c.[-276C>A];[=]	Local homopolymer	Adjacent/Internal to homopolymer
34	10183256	c.-277_-276insA	Local homopolymer	Internal to homopolymer
35	10184026	c.340+154_340+155insT;c.340+155_340+156delTT	Local homopolymer	Internal to homopolymer

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
36	10184026	c.340+154_340+155insTT;c.340+155delT	Local homopolymer	Internal to homopolymer
37	10184026	c.340+154_340+155insT	Local homopolymer	Internal to homopolymer
38	10184026	c.340+154_340+155insTT	Local homopolymer	Internal to homopolymer
39	10184026	c.340+154_340+155insT;c.340+155delT	Local homopolymer	Internal to homopolymer
40	10184026	c.340+154_340+155insT;c.340+155_340+157delTTT	Local homopolymer	Internal to homopolymer
41	10184044	c.340+173_340+174delTT	Artefact	Low percentage reads, all samples within 1 or more runs
42	10184045	c.340+174delT	Local homopolymer	Internal to homopolymer
43	10184046	c.340+174_340+175insT	Local homopolymer	Internal to homopolymer
44	10184046	c.340+174_340+175insTT	Local homopolymer	Internal to homopolymer
45	10184046	c.340+174_340+175insTT;c.[340+175C>T];[=]	Local homopolymer	Adjacent/Internal to homopolymer
46	10184046	c.340+174_340+175insT;c.[340+175C>T];[=]	Local homopolymer	Adjacent/Internal to homopolymer
47	10184046	c.[340+175C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
48	10184047	c.[340+176T>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
49	10184049	c.[340+178A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
50	10184055	c.[340+184A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
51	10184059	c.[340+188T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
52	10184063	c.[340+192T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
53	10184100	c.[340+229A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
54	10184123	c.[340+252G>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
55	10184126	c.[340+255T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
56	10184182	c.[340+311A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
57	10184227	c.[340+356A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
58	10184259	c.[340+388A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
59	10184263	c.[340+392C>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
60	10184268	c.[340+397T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
61	10184802	c.340+930_340+931insT;c.340+931_340+932delTT	Local homopolymer	Internal to homopolymer
62	10184802	c.340+930_340+931insT;c.340+931delT	Local homopolymer	Internal to homopolymer
63	10184815	c.340+944_340+945delTT	Local homopolymer	Internal to homopolymer
64	10184816	c.340+945delT	Local homopolymer	Internal to homopolymer
65	10184817	c.340+945_340+946insT	Local homopolymer	Internal to homopolymer
66	10184982	c.340+1111delT	Local homopolymer	Internal to homopolymer
67	10184982	c.340+1110_340+1111insT;c.340+1111delT	Local homopolymer	Internal to homopolymer
68	10184991	c.340+1120delT	Local homopolymer	Internal to homopolymer
69	10184992	c.340+1120_340+1121insT	Local homopolymer	Internal to homopolymer
70	10185056	c.[340+1185A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
71	10185990	c.340+2118_340+2119insA;c.340+2119delA	Local homopolymer	Internal to homopolymer

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
72	10185990	c.340+2118_340+2119insA	Local homopolymer	Internal to homopolymer
73	10186003	c.340+2132delA	Local homopolymer	Internal to homopolymer
74	10186004	c.340+2132_340+2133insA	Local homopolymer	Internal to homopolymer
75	10186181	c.341-2018_341-2017insT;c.341-2017_341-2015delTTT	Local homopolymer	Internal to homopolymer
76	10186181	c.341-2018_341-2017insT;c.341-2017_341-2016delTT	Local homopolymer	Internal to homopolymer
77	10186181	c.341-2018_341-2017insT;c.341-2017delT	Local homopolymer	Internal to homopolymer
78	10186197	c.341-2001_341-2000delTT	Local homopolymer	Internal to homopolymer
79	10186198	c.341-2000delT	Local homopolymer	Internal to homopolymer
80	10186199	c.341-2000_341-1999insT	Local homopolymer	Internal to homopolymer
81	10187161	c.[341-1037G>A];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
82	10187167	c.[341-1031A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
83	10187172	c.[341-1026A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
84	10187175	c.[341-1023A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
85	10187299	c.[341-899A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
86	10187346	c.[341-852T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
87	10187368	c.[341-830C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
88	10187390	c.341-809_341-808insA;c.341-808_341-807delAA	Local homopolymer	Internal to homopolymer
89	10187390	c.341-809_341-808insA;c.341-808delA	Local homopolymer	Internal to homopolymer
90	10187402	c.341-796_341-795delAA	Local homopolymer	Internal to homopolymer
91	10187403	c.341-795delA	Local homopolymer	Internal to homopolymer
92	10187404	c.341-795_341-794insA	Local homopolymer	Internal to homopolymer
93	10187858	c.341-341_341-340insT;c.341-340_341-339delTT	Local homopolymer	Internal to homopolymer
94	10187858	c.341-341_341-340insT;c.341-340_341-338delTTT	Local homopolymer	Internal to homopolymer
95	10187872	c.341-326_341-324delTTT	Local homopolymer	Internal to homopolymer

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
96	10187873	c.341-325_341-324delTT	Local homopolymer	Internal to homopolymer
97	10187875	c.341-324_341-323insT	Local homopolymer	Internal to homopolymer
98	10187883	c.[341-315A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
99	10187897	c.[341-301A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
100	10188986	c.[463+666A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
101	10189013	c.[463+693A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
102	10189044	c.463+723_463+724insT;c.463+724_463+725delTT	Local homopolymer	Internal to homopolymer
103	10189044	c.463+724delT	Local homopolymer	Internal to homopolymer
104	10189044	c.463+723_463+724insT;c.463+724delT	Local homopolymer	Internal to homopolymer
105	10189056	c.463+736_463+737delTT	Local homopolymer	Internal to homopolymer
106	10189056	c.[463+736T>A];[=]	Local homopolymer	Internal to homopolymer
107	10189057	c.[463+737T>A];[=]	Local homopolymer	Adjacent to homopolymer

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
108	10189058	c.463+737_463+738insTA	Local homopolymer	Internal to homopolymer
109	10189058	c.463+737_463+738insA	Local homopolymer	Internal to homopolymer
110	10189059	c.463+738_463+739insA	Local homopolymer	Internal to homopolymer
111	10189296	c.[463+976A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
112	10189351	c.463+1030_463+1031insT;c.463+1031_463+1032delTT	Local homopolymer	Internal to homopolymer
113	10189364	c.463+1044_463+1045delTT	Local homopolymer	Internal to homopolymer
114	10189366	c.463+1045_463+1046insT	Local homopolymer	Internal to homopolymer
115	10189554	c.[463+1234T>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
116	10189560	c.[463+1240G>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
117	10189566	c.[463+1246A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
118	10189569	c.[463+1249A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
119	10189571	c.[463+1251T>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
120	10189587	c.[463+1267A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
121	10189611	c.[463+1291A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
122	10189788	c.463+1467_463+1468insA;c.463+1468delA	Local homopolymer	Internal to homopolymer
123	10189797	c.463+1477delA	Local homopolymer	Internal to homopolymer
124	10189798	c.463+1477_463+1478insA	Local homopolymer	Internal to homopolymer
125	10189999	c.[464-1472C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
126	10190000	c.464-1472_464-1471insT;c.464-1471_464-1470delTT	Local homopolymer	Internal to homopolymer
127	10190000	c.464-1472_464-1471insT;c.464-1471_464-1469delTTT	Local homopolymer	Internal to homopolymer
128	10190014	c.464-1457_464-1455delTTT	Local homopolymer	Internal to homopolymer
129	10190015	c.464-1456_464-1455delTT	Local homopolymer	Internal to homopolymer
130	10190016	c.464-1455delT	Local homopolymer	Internal to homopolymer
131	10190017	c.464-1455_464-1454insT	Local homopolymer	Internal to homopolymer

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
132	10190063	c.[464-1408A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
133	10190065	c.[464-1406C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
134	10190073	c.[464-1398A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
135	10190092	c.[464-1379G>A];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
136	10190095	c.[464-1376T>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
137	10190099	c.[464-1372C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
138	10190124	c.[464-1347A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
139	10190153	c.[464-1318T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
140	10190153	c.[464-1318T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
141	10190172	c.[464-1299G>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
142	10190174	c.[464-1297A>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
143	10190179	c.[464-1292T>A];[=]	Artefact	Low percentage reads, all samples within 1 or more runs

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
144	10190187	c.[464-1284A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
145	10190192	c.[464-1279G>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
146	10190197	c.[464-1274A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
147	10190200	c.[464-1271G>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
148	10190202	c.[464-1269T>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
149	10190209	c.[464-1262G>A];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
150	10190224	c.[464-1247T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
151	10190229	c.[464-1242A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
152	10190230	c.[464-1241C>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
153	10190234	c.[464-1237G>A];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
154	10190238	c.[464-1233T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
155	10190240	c.[464-1231C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
156	10190364	c.[464-1107A>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
157	10190367	c.464-1105_464-1104insT;c.464-1104delT	Local homopolymer	Internal to homopolymer
158	10190367	c.464-1105_464-1104insT;c.464-1104_464-1103delTT	Local homopolymer	Internal to homopolymer
159	10190379	c.464-1092_464-1091delTT	Local homopolymer	Internal to homopolymer
160	10190380	c.464-1091delT	Local homopolymer	Internal to homopolymer
161	10190381	c.464-1091_464-1090insT	Local homopolymer	Internal to homopolymer
162	10190779	c.464-693_464-692insT;c.464-692delT	Local homopolymer	Internal to homopolymer
163	10190779	c.464-693_464-692insT;c.464-692_464-691delTT	Local homopolymer	Internal to homopolymer
164	10190789	c.464-682_464-681delTT	Local homopolymer	Internal to homopolymer
165	10190790	c.464-681delT	Local homopolymer	Internal to homopolymer
166	10190791	c.464-681_464-680insT	Local homopolymer	Internal to homopolymer
167	10193057	c.*1408delT	Local homopolymer	Internal to homopolymer

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
168	10193065	c.*1416delT	Local homopolymer	Internal to homopolymer
169	10193452	c.[*1803C>A];[=]	Local homopolymer	Internal to homopolymer
170	10193745	c.*2096_*2099delAAAA	Local homopolymer	Internal to homopolymer
171	10193745	c.*2096_*2098delAAA	Local homopolymer	Internal to homopolymer
172	10193745	c.*2095_*2096insA;c.*2096_*2098delAAA	Local homopolymer	Internal to homopolymer
173	10193763	c.*2114_*2117delAAAA	Local homopolymer	Internal to homopolymer
174	10193767	c.*2117_*2118insA	Local homopolymer	Internal to homopolymer
175	10193816	c.[*2167T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
176	10193827	c.[*2178A>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
177	10194054	c.*2404_*2405insA;c.*2405_*2406delAA	Local homopolymer	Internal to homopolymer
178	10194067	c.*2418_*2419delAA	Local homopolymer	Internal to homopolymer
179	10194069	c.*2419_*2420insA	Local homopolymer	Internal to homopolymer

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
180	10194137	c.[*2488C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
181	10194140	c.[*2491C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
182	10194151	c.[*2502A>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
183	10194160	c.[*2511A>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
184	10194163	c.[*2514G>T];[=]	Local homopolymer	Internal to homopolymer
185	10194165	c.[*2516C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
186	10194166	c.*2517_*2521delTTTTT	Local homopolymer	Internal to homopolymer
187	10194166	c.*2517_*2520delTTTT	Local homopolymer	Internal to homopolymer
188	10194166	c.*2517_*2519delTTT	Local homopolymer	Internal to homopolymer
189	10194186	c.*2537_*2545delTTTTTTTTT	Local homopolymer	Internal to homopolymer
190	10194187	c.*2538_*2545delTTTTTTTT	Local homopolymer	Internal to homopolymer
191	10194188	c.*2539_*2545delTTTTTTT	Local homopolymer	Internal to homopolymer

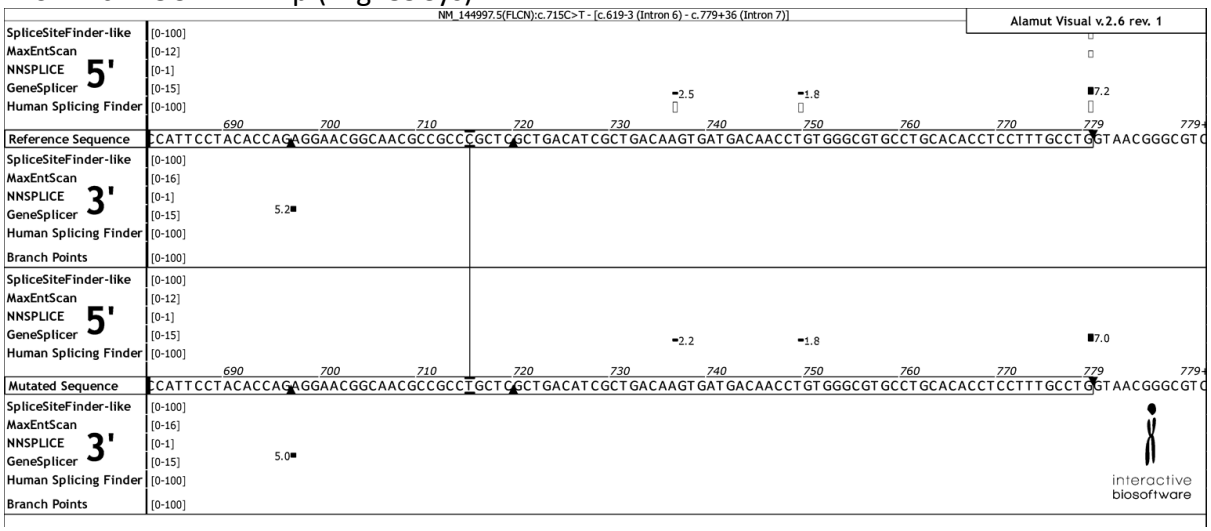
Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
192	10194195	c.*2546delG	Local homopolymer	Internal to homopolymer
193	10194197	c.[*2548G>A];[=]	Local homopolymer	Internal to homopolymer
194	10194409	c.[*2760T>C];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
195	10194726	c.[*3077C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
196	10194727	c.[*3078T>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
197	10194731	c.[*3082C>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
198	10194735	c.[*3086T>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
199	10194743	c.[*3094G>A];[=]	Local homopolymer	Adjacent to homopolymer
200	10194827	c.[*3178T>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
201	10194830	c.[*3181T>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
202	10194962	c.*3312_*3313insA;c.*3313delA	Local homopolymer	Internal to homopolymer
203	10194972	c.*3323delA	Local homopolymer	Internal to homopolymer

Number	Chromosome position - Start	Nucleotide change	Likely cause of false call	Details
204	10194973	c.*3323_*3324insA	Local homopolymer	Internal to homopolymer
205	10195007	c.[*3358C>G];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
206	10195026	c.[*3377T>A];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
207	10195032	c.[*3383G>T];[=]	Artefact	Low percentage reads, all samples within 1 or more runs
208	10195133	c.*3483_*3484insA;c.*3484delA	Local homopolymer	Internal to homopolymer
209	10195142	c.*3493delA	Local homopolymer	Internal to homopolymer
210	10195143	c.*3493_*3494insA	Local homopolymer	Internal to homopolymer
211	10195307	c.*3658_*3659delTT	Local homopolymer	Internal to homopolymer
212	10195309	c.*3659_*3660insT	Local homopolymer	Internal to homopolymer

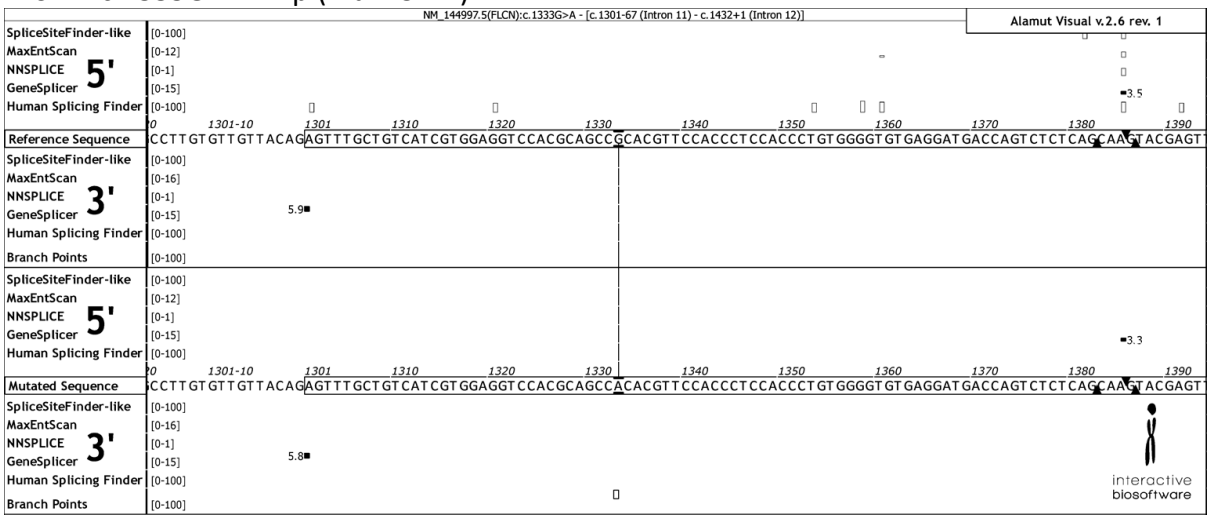
7.3 Chapter 5 Appendices

Table 7.4 Tabulated view of splicing images using Alamut as an interface to interrogate five splicing algorithms

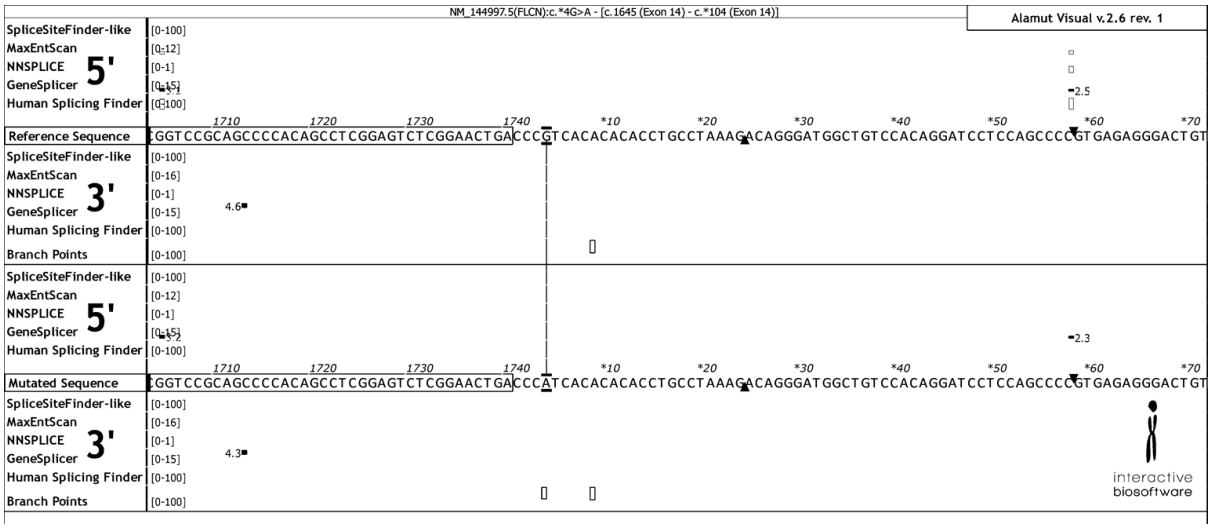
FLCN c.715C>T p.(Arg239Cys)



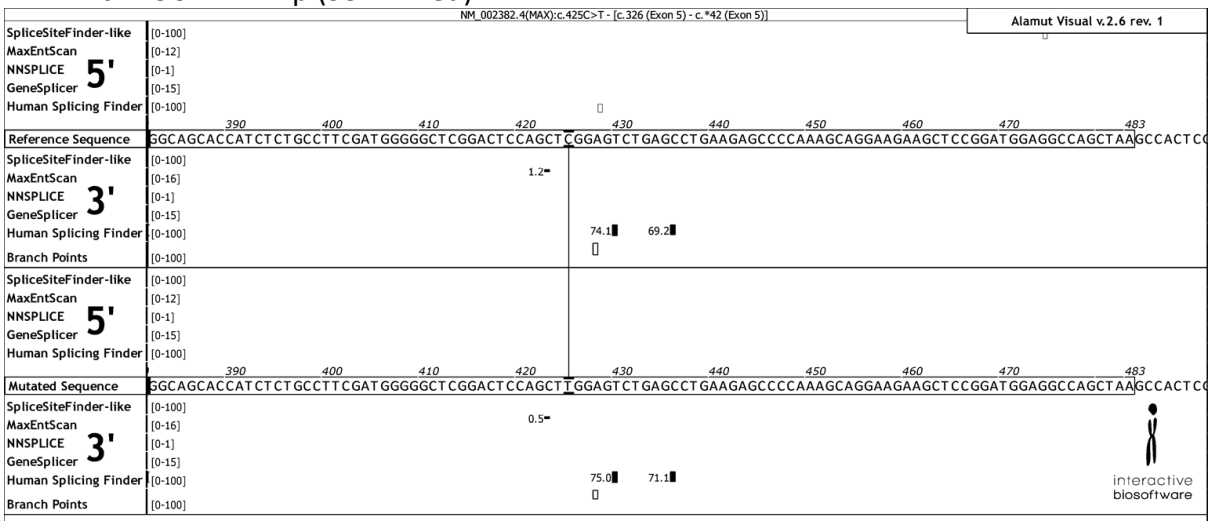
FLCN c.1333G>A p.(Ala445Thr)



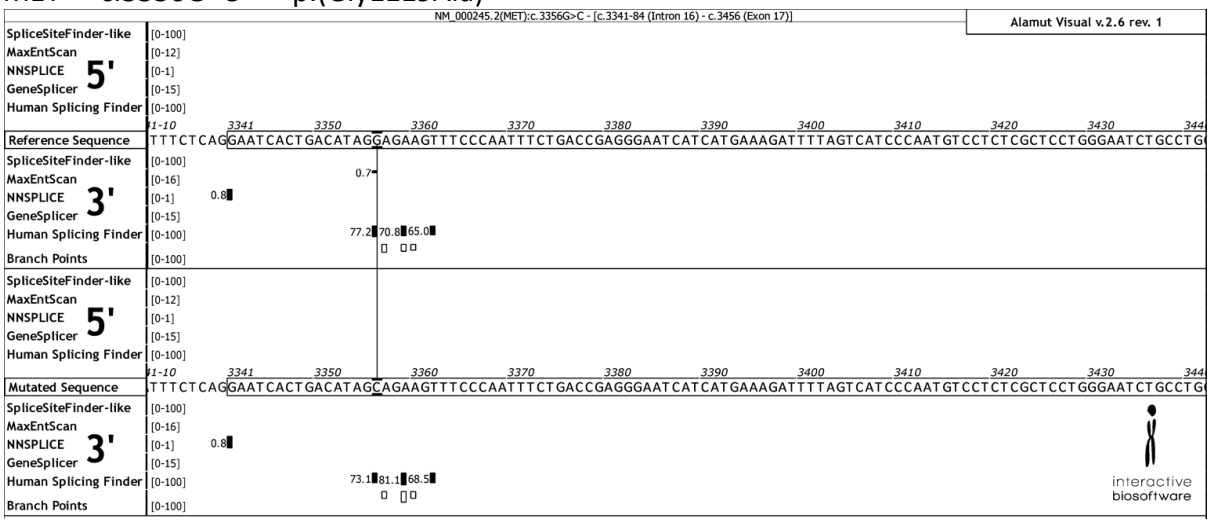
FLCN c.*4G>A



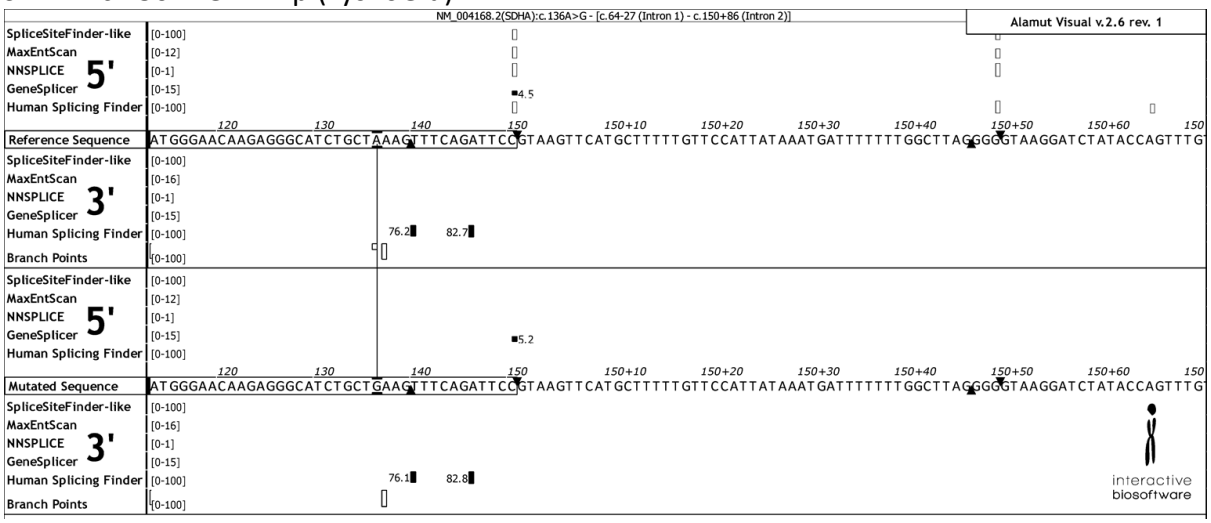
MAX c.425C>T p.(Ser142Leu)



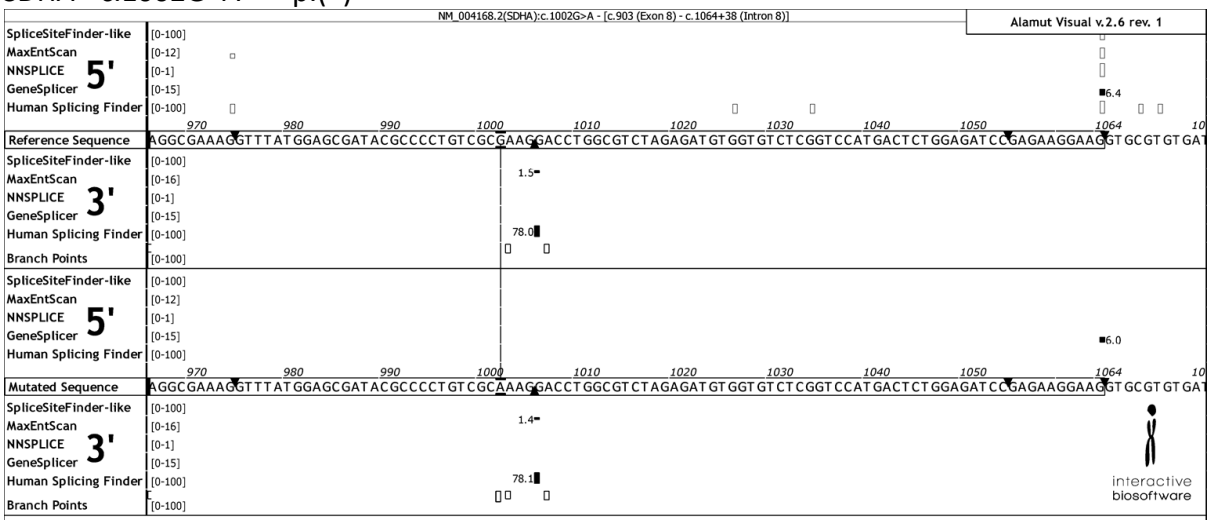
MET c.3356G>C p.(Gly1119Ala)



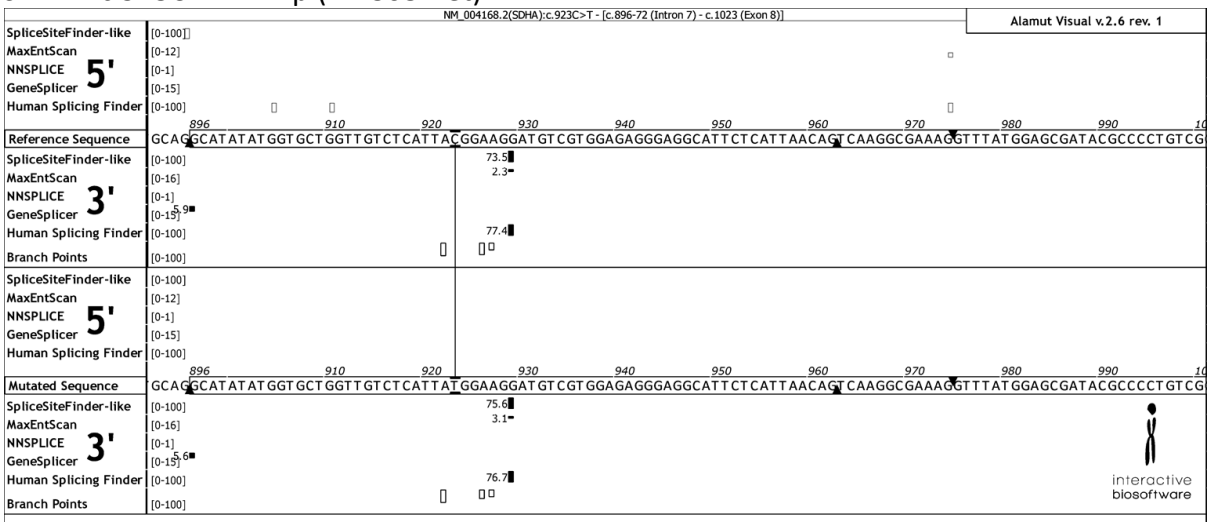
SDHA c.136A>G p.(Lys46Glu)



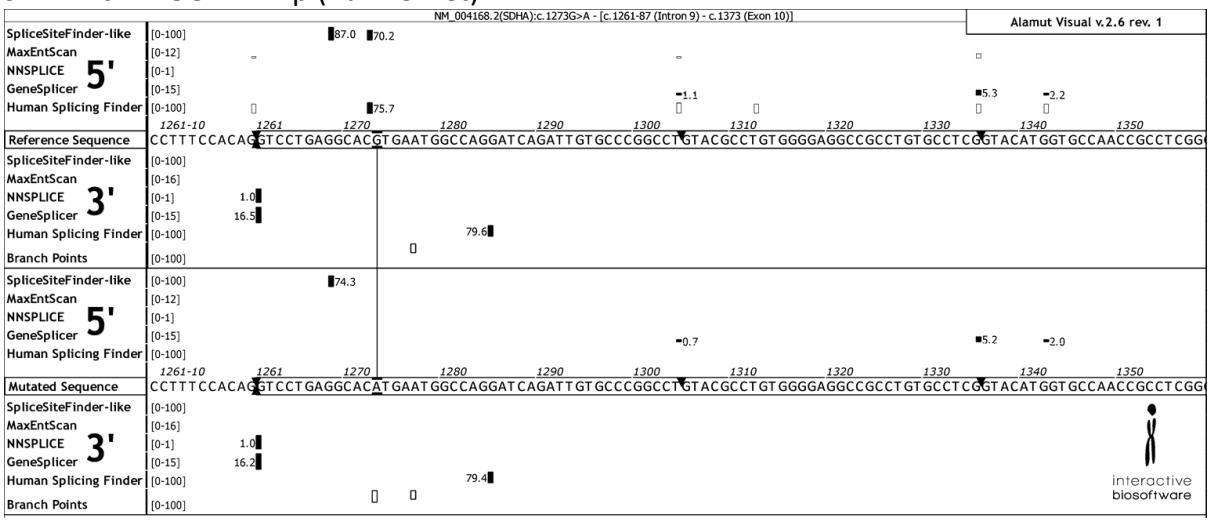
SDHA c.1002G>A p.(=)



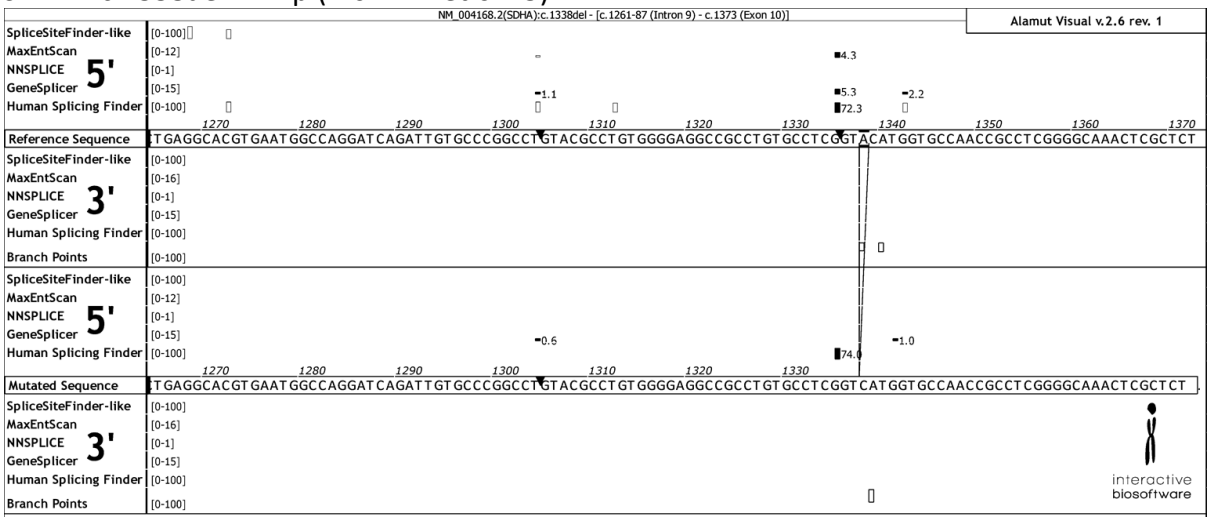
SDHA c.923C>T p.(Thr308Met)



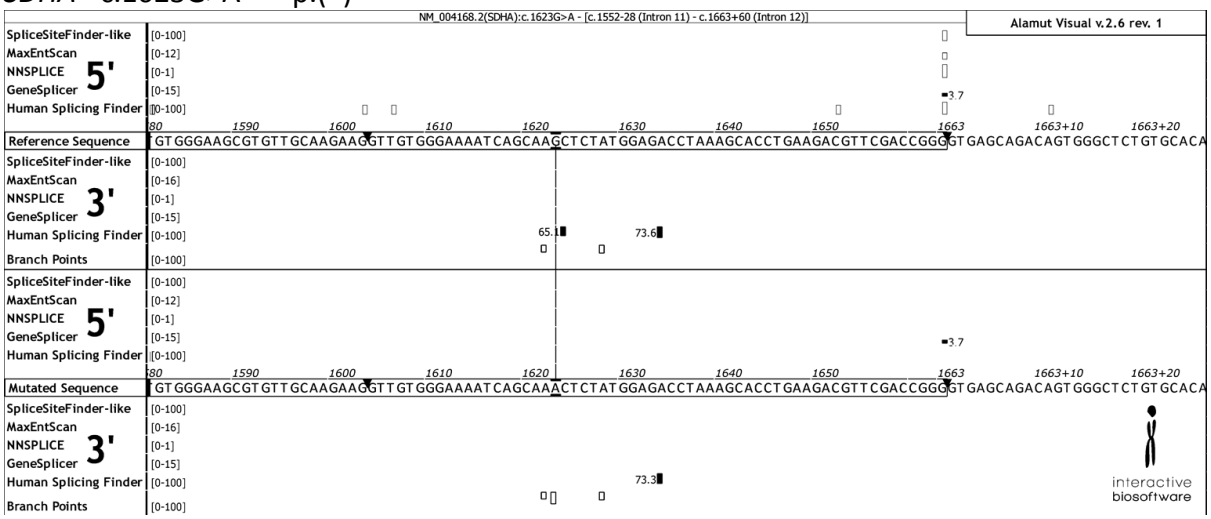
SDHA c.1273G>A p.(Val425Met)



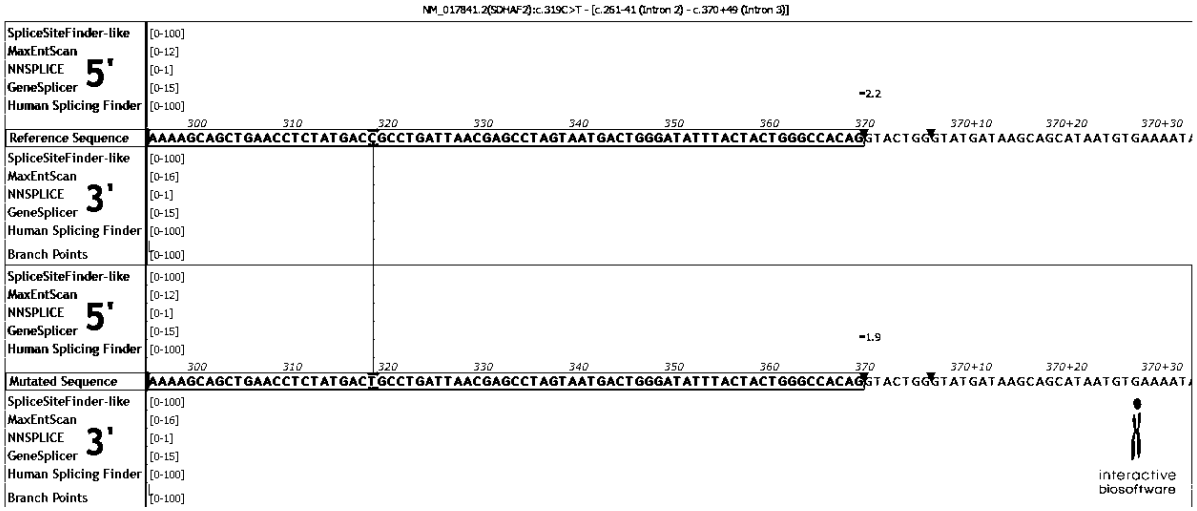
SDHA c.1338delA p.(His447Metfs*23)



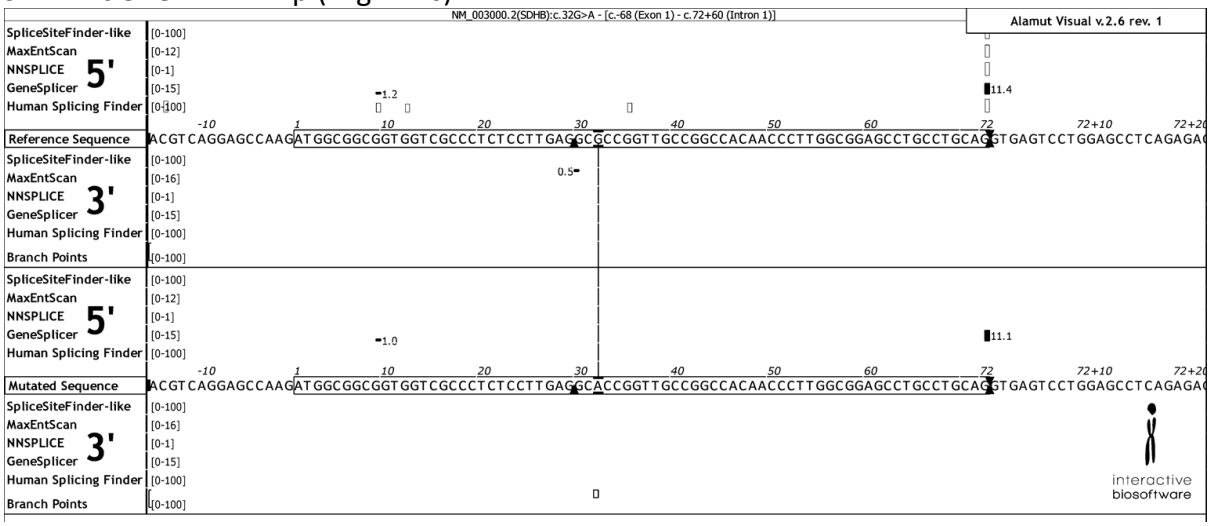
SDHA c.1623G>A p.(=)



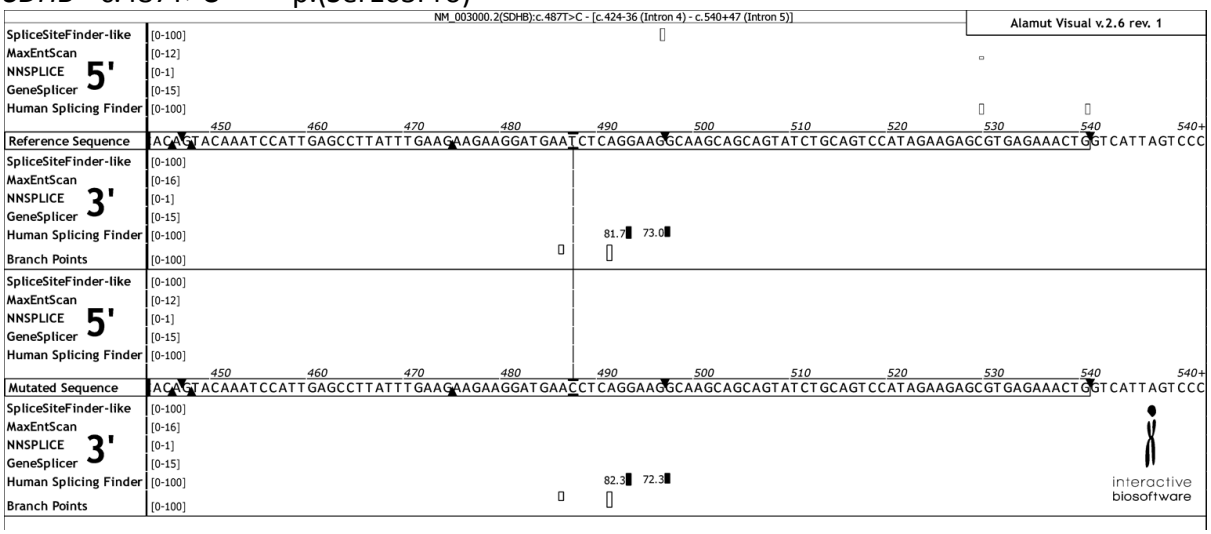
SDHAF2 c.319C>T p.(Arg107Cys)



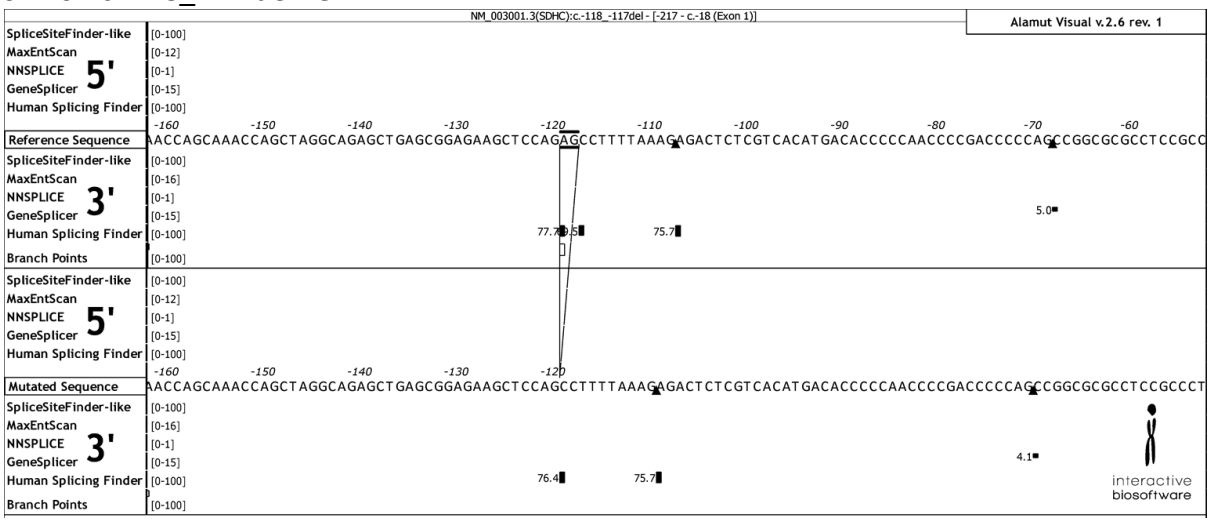
SDHB c.32G>A p.(Arg11His)



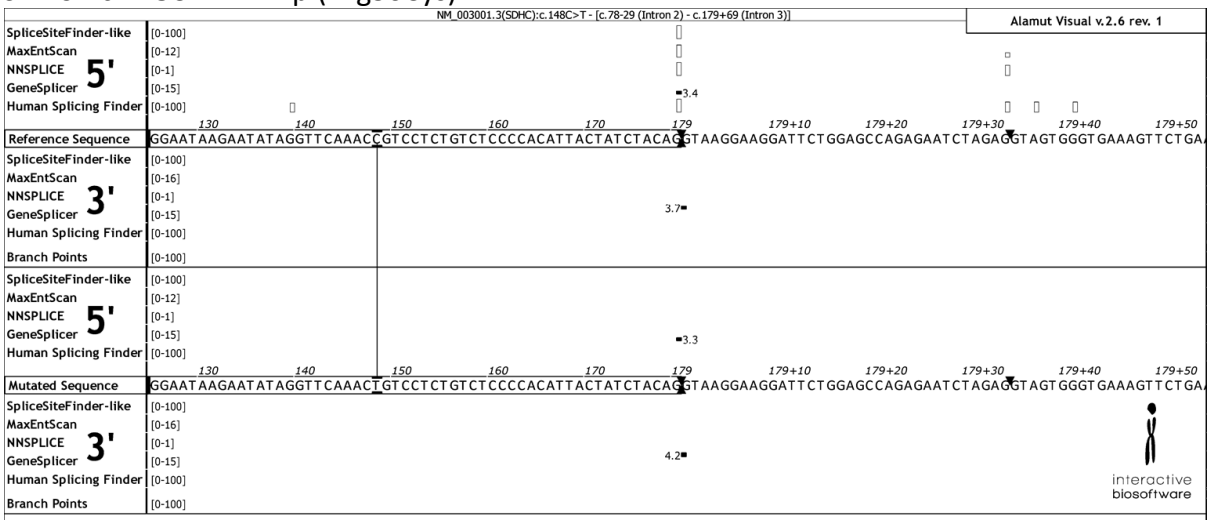
SDHB c.487T>C p.(Ser163Pro)



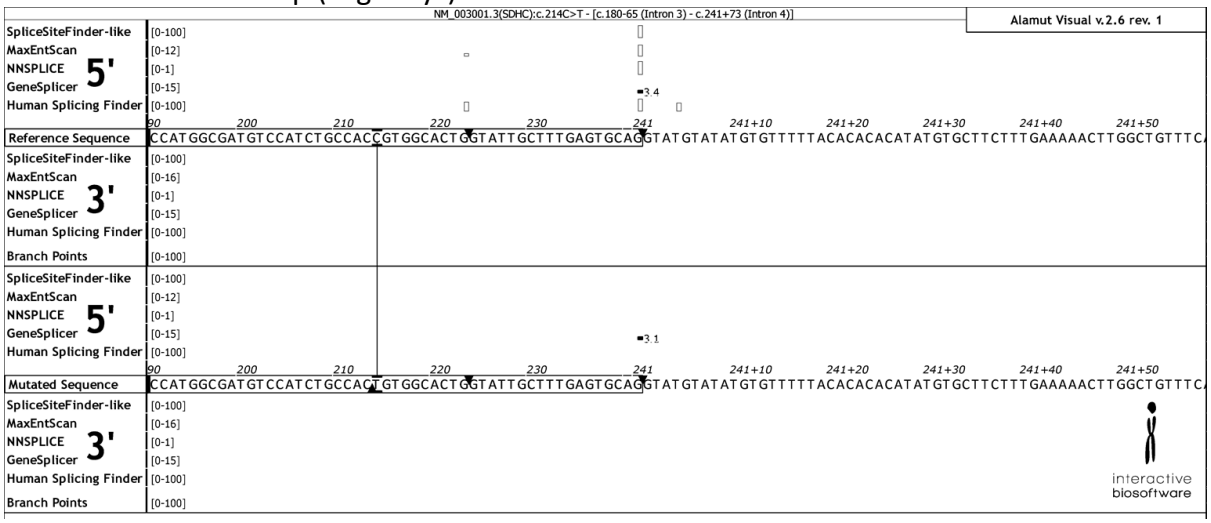
SDHC c.-118_-117delAG



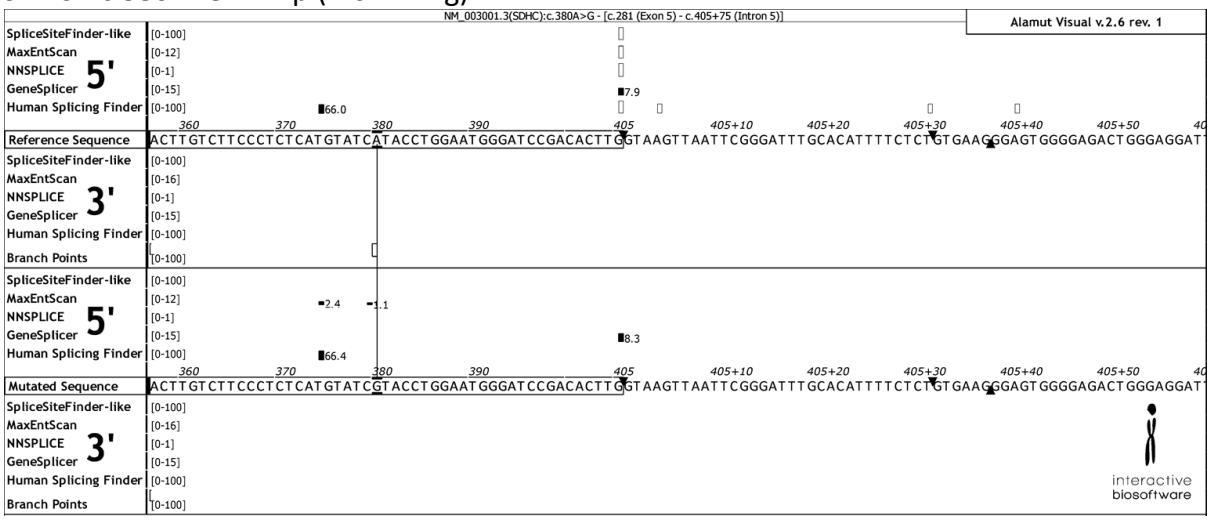
SDHC c.148C>T p.(Arg50Cys)



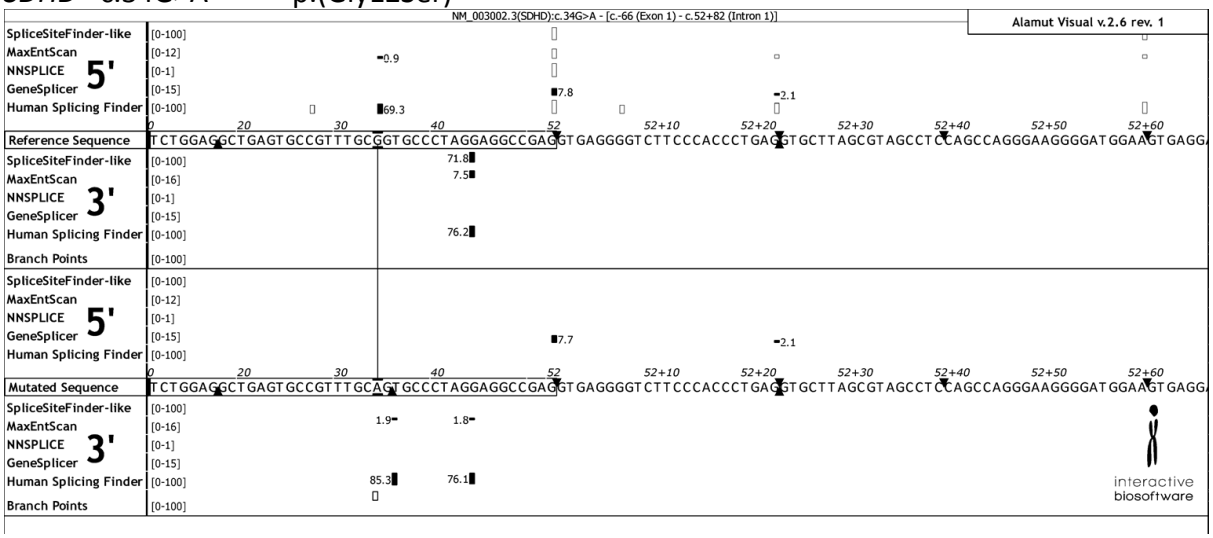
SDHC c.214C>T p.(Arg72Cys)



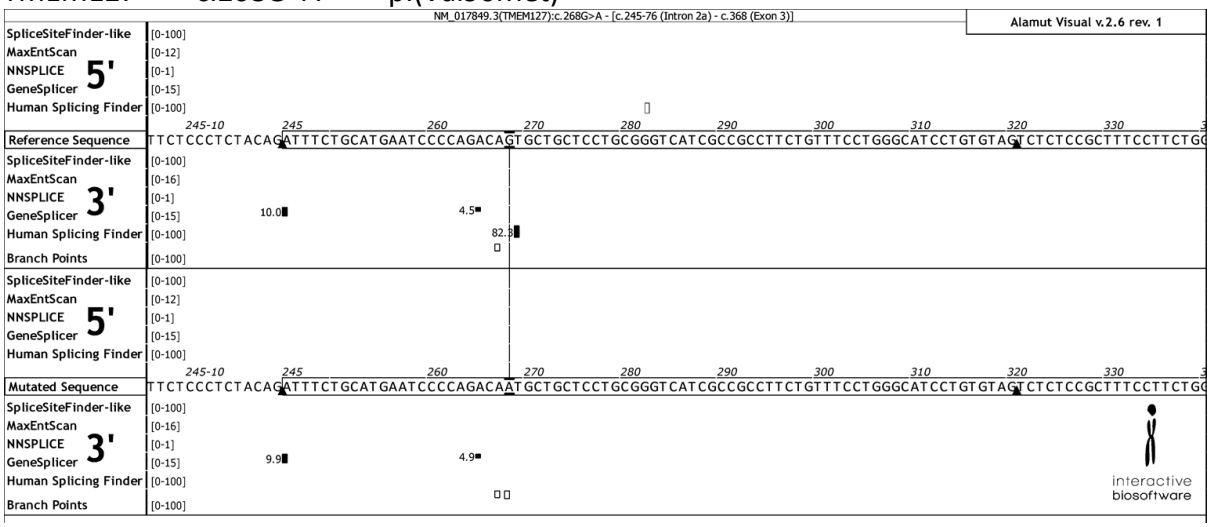
SDHC c.380A>G p.(His127Arg)



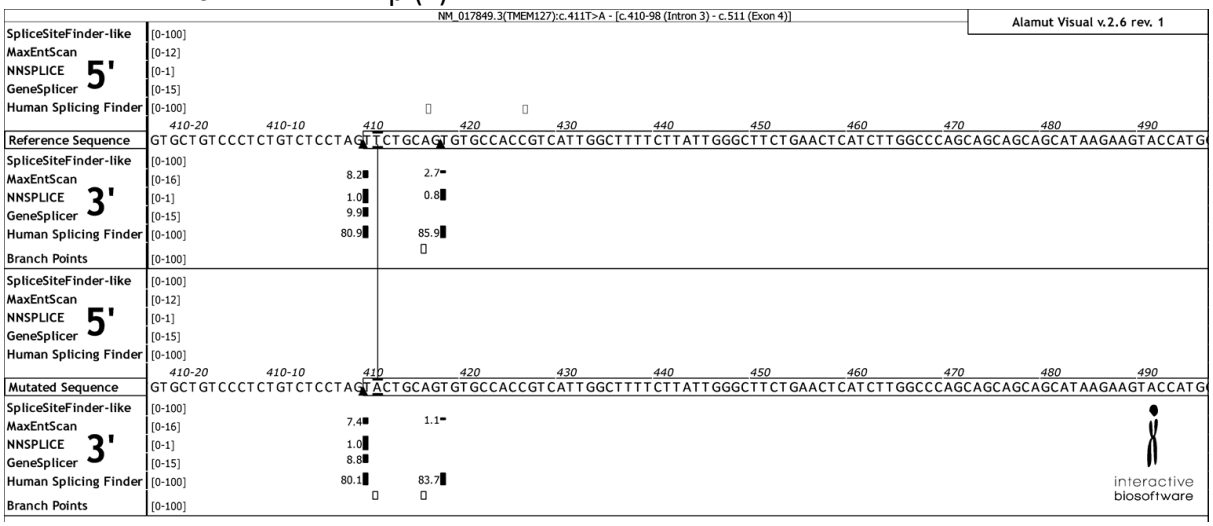
SDHD c.34G>A p.(Gly12Ser)



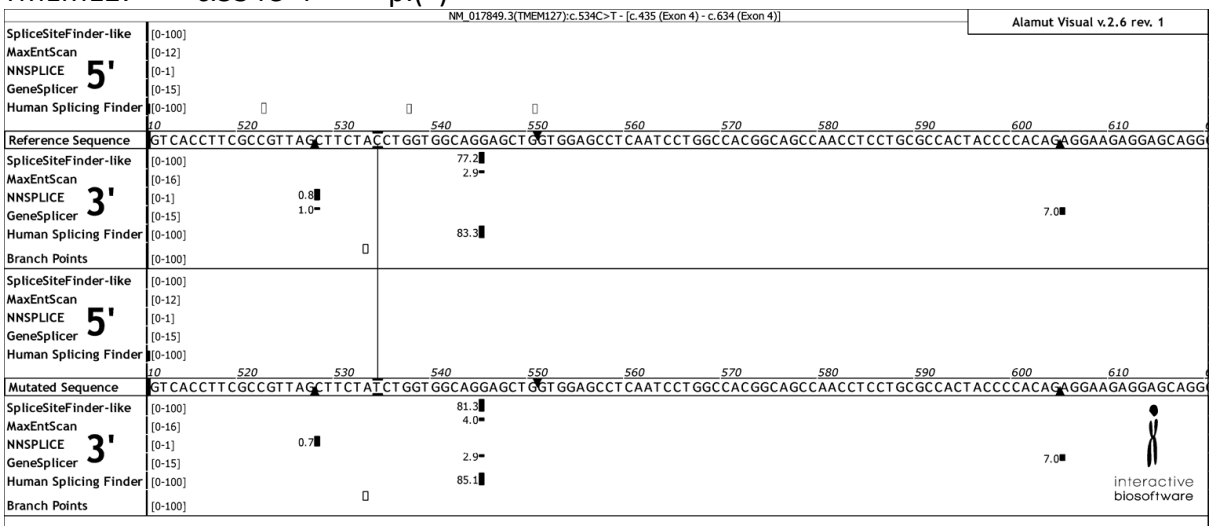
TMEM127 c.268G>A p.(Val90Met)



TMEM127 c.411T>A p.(=)



TMEM127 c.534C>T p.(=)



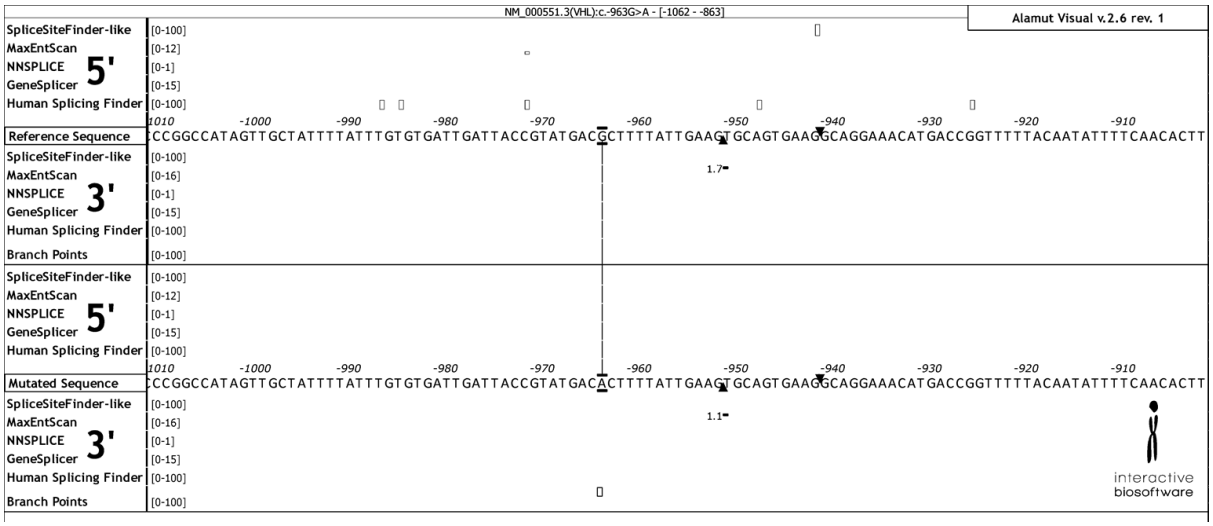
VHL c.-3933C>T

No prediction available

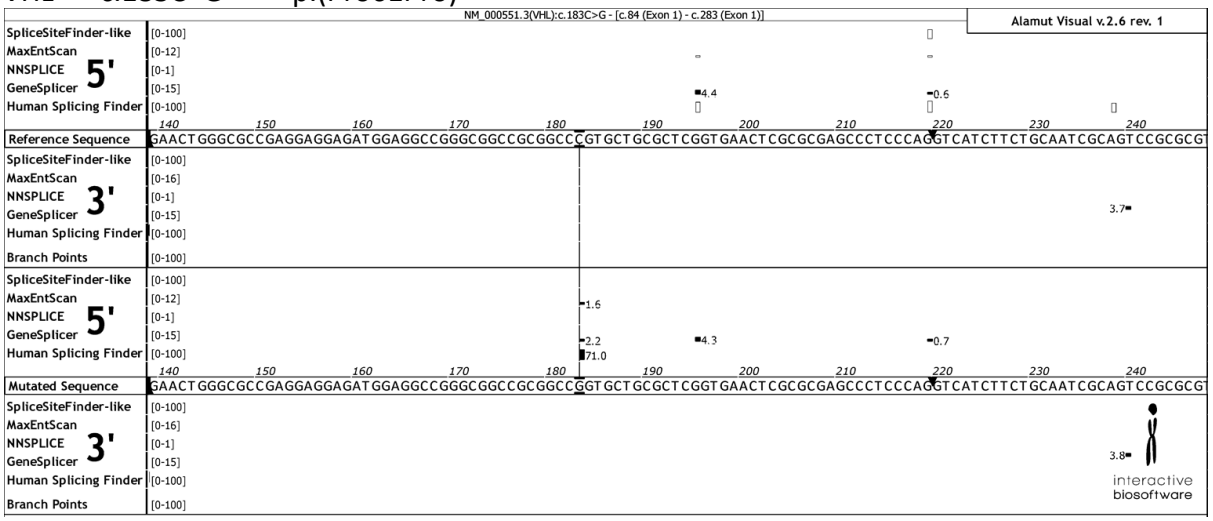
VHL c.-3197_-3195delCTC

No prediction available

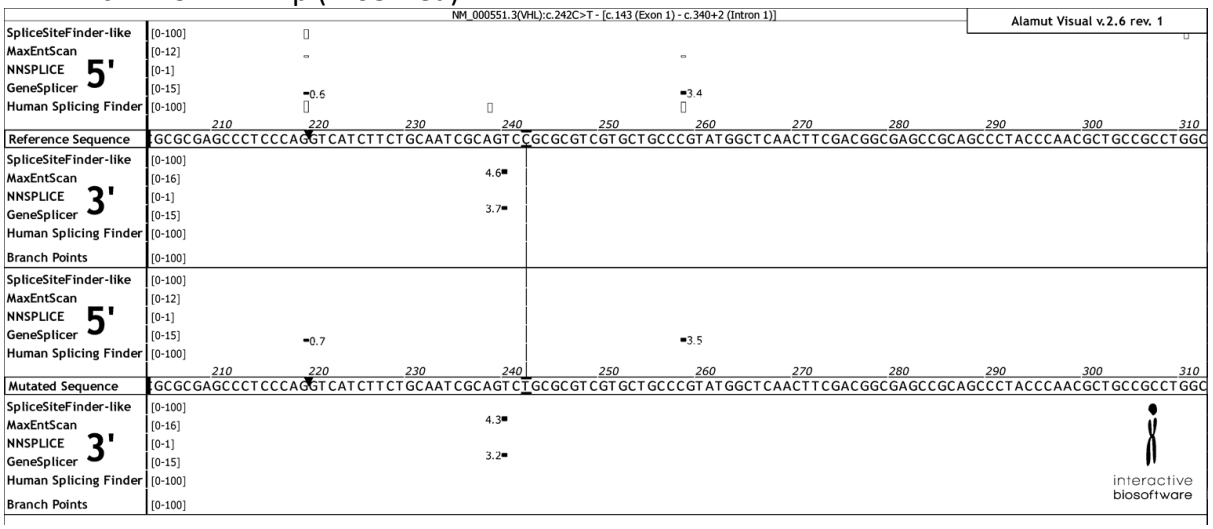
VHL c.-963G>A



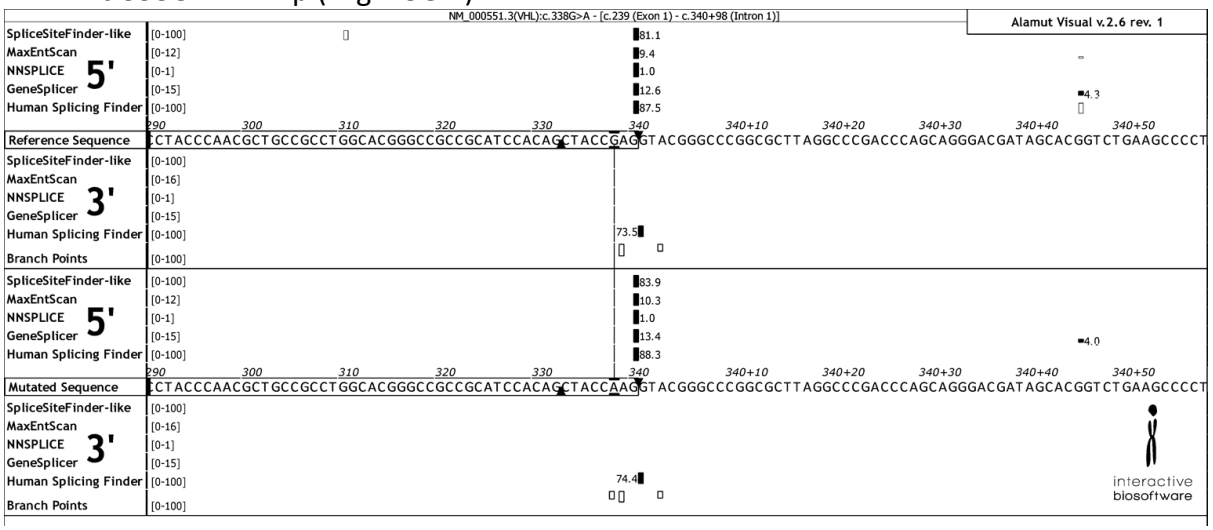
VHL c.183C>G p.(Pro61Pro)



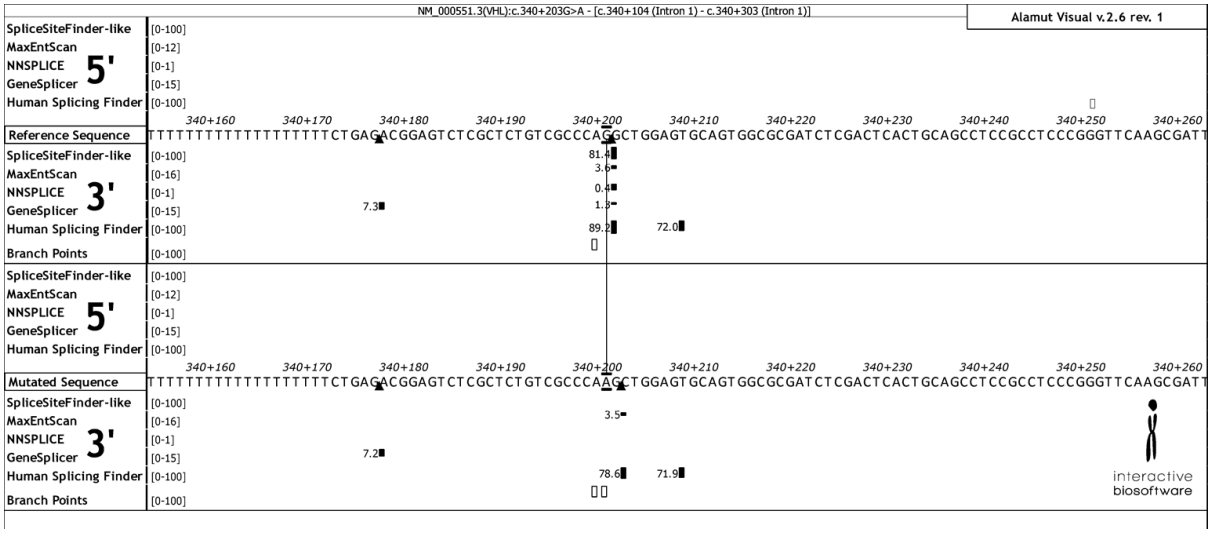
VHL c.242C>T p.(Pro81Leu)



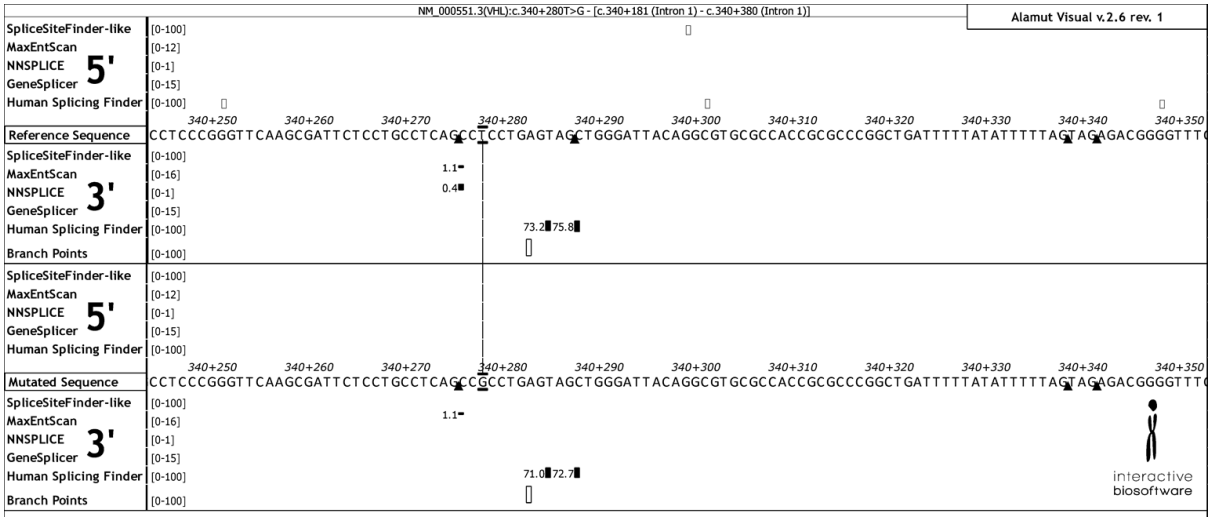
VHL c.338G>A p.(Arg113Gln)



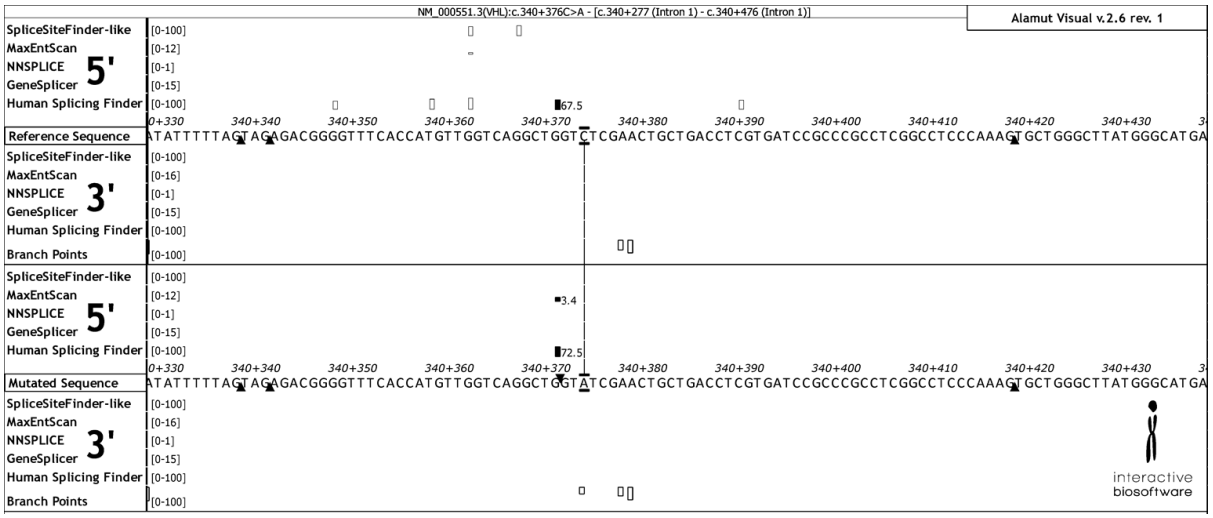
VHL c.340+203G>A



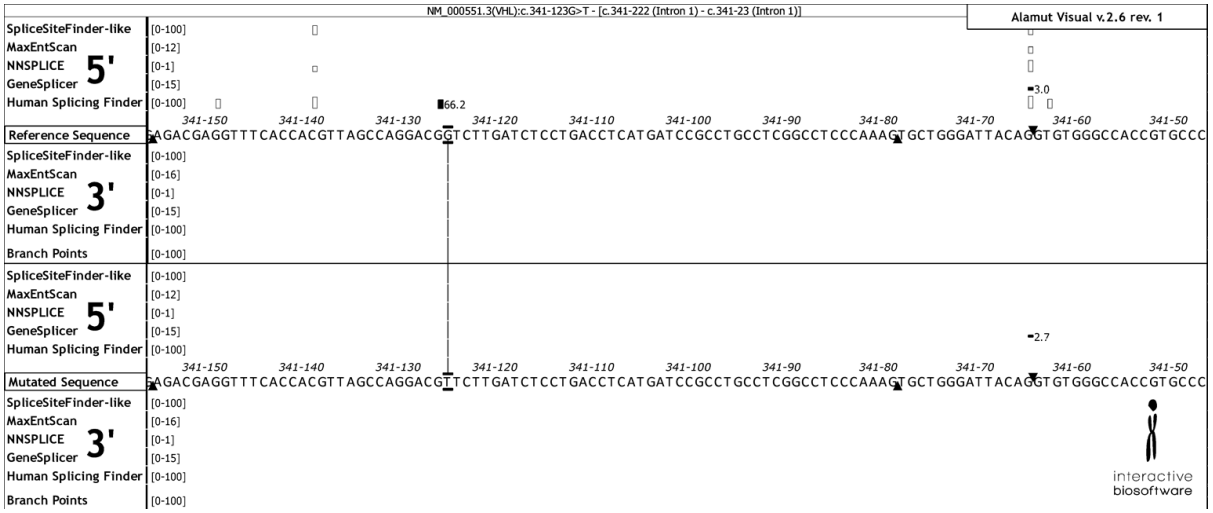
VHL c.340+280T>G



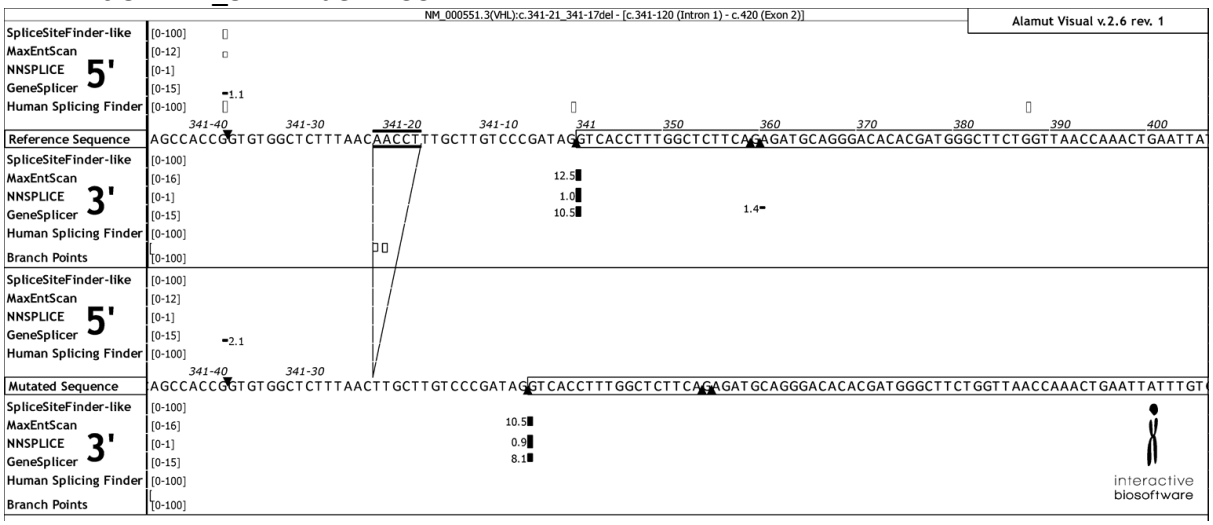
VHL c.340+376C>A



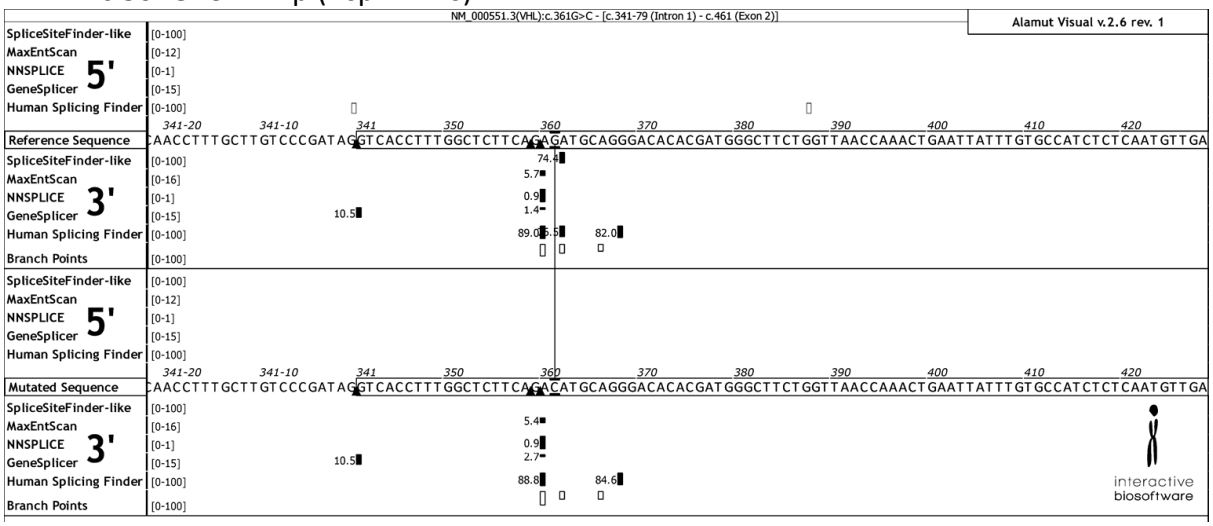
VHL c.341-123G>T



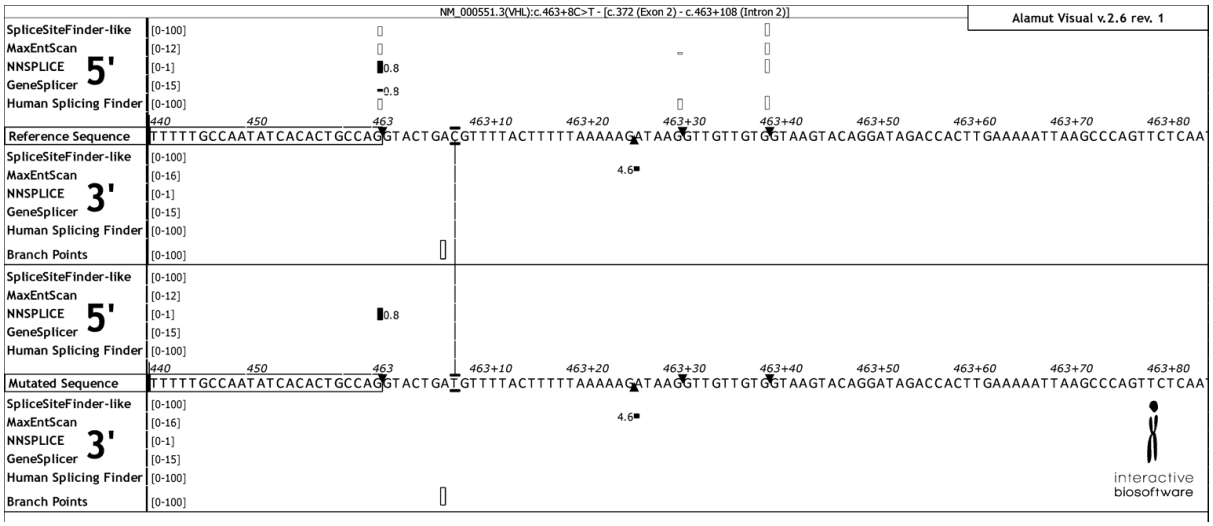
VHL c.341-21_341-17delAACCT



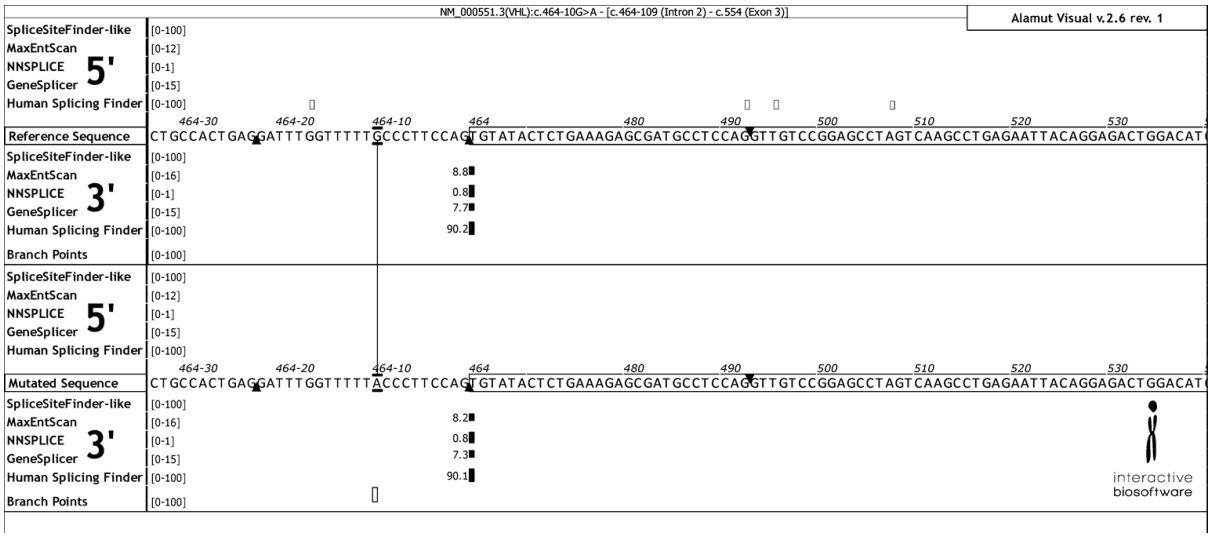
VHL c.361G>C p.(Asp121His)



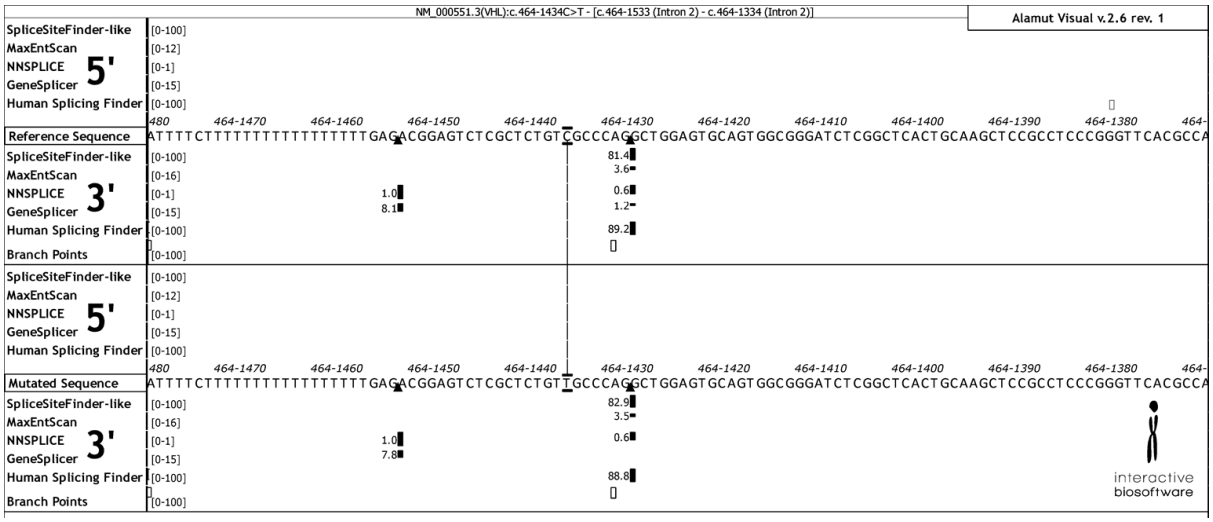
VHL c.463+8C>T



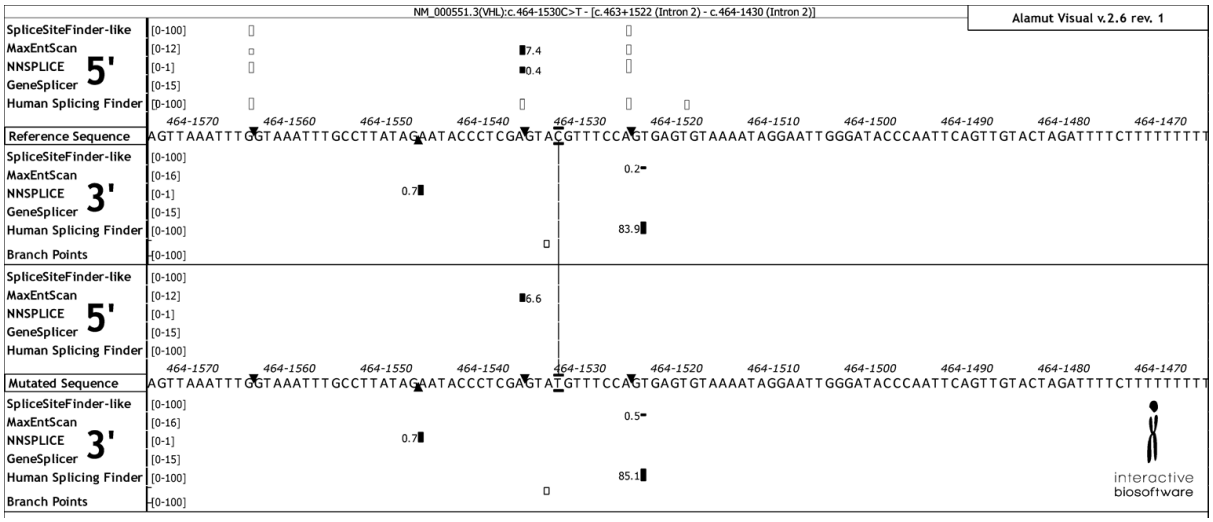
VHL c.464-10G>A



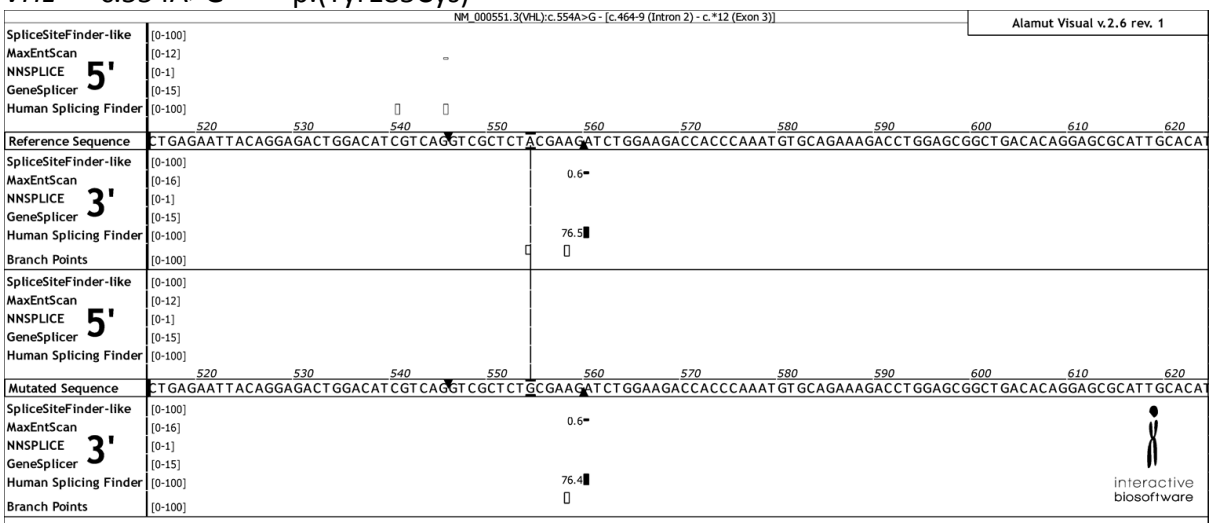
VHL c.464-1434C>T



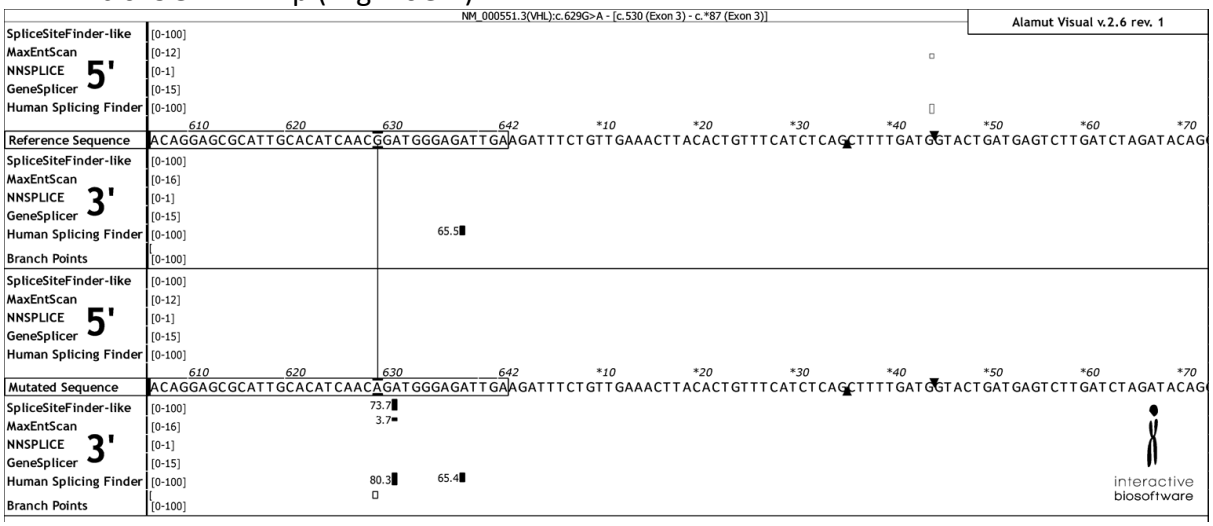
VHL c.464-1530C>T



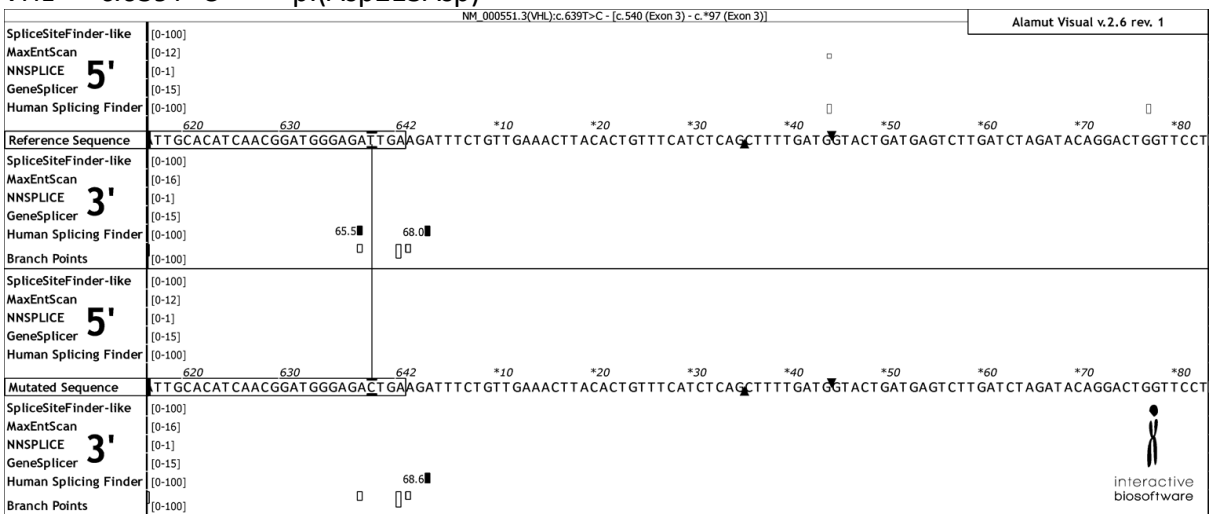
VHL c.554A>G p.(Tyr185Cys)



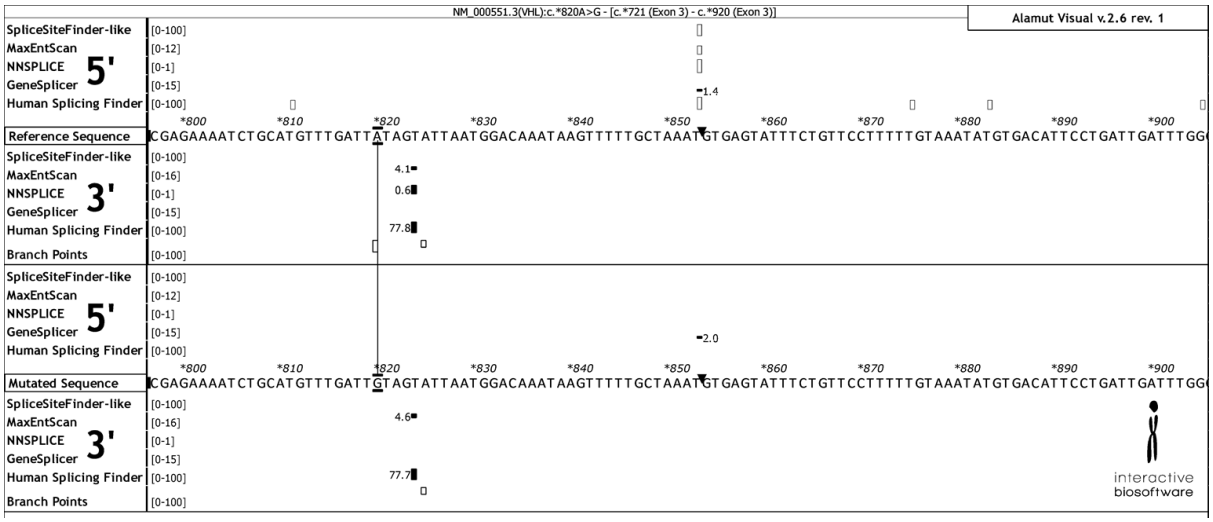
VHL c.629G>A p.(Arg210Gln)



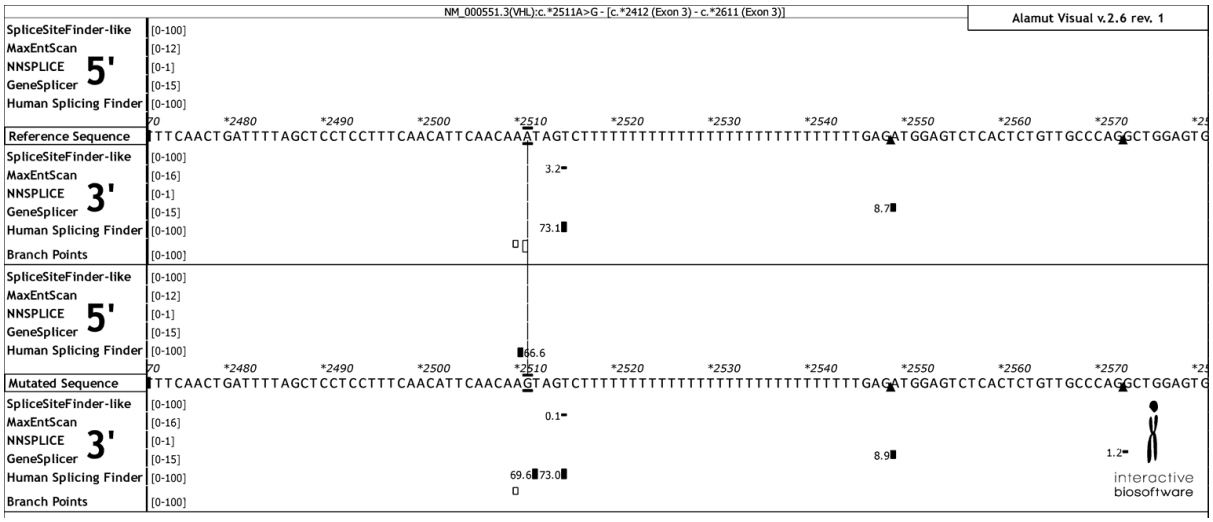
VHL c.639T>C p.(Asp213Asp)



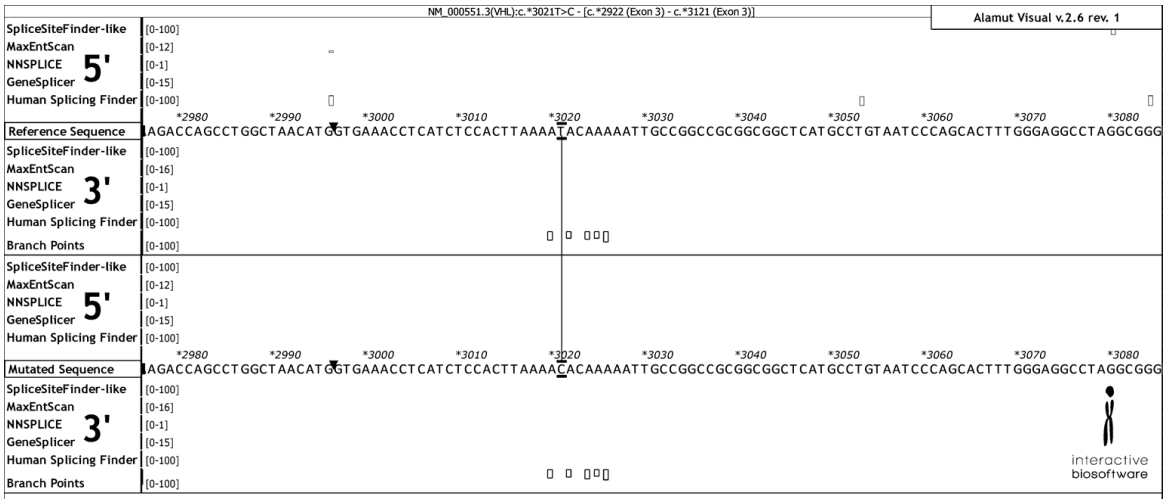
VHL c.*820A>G



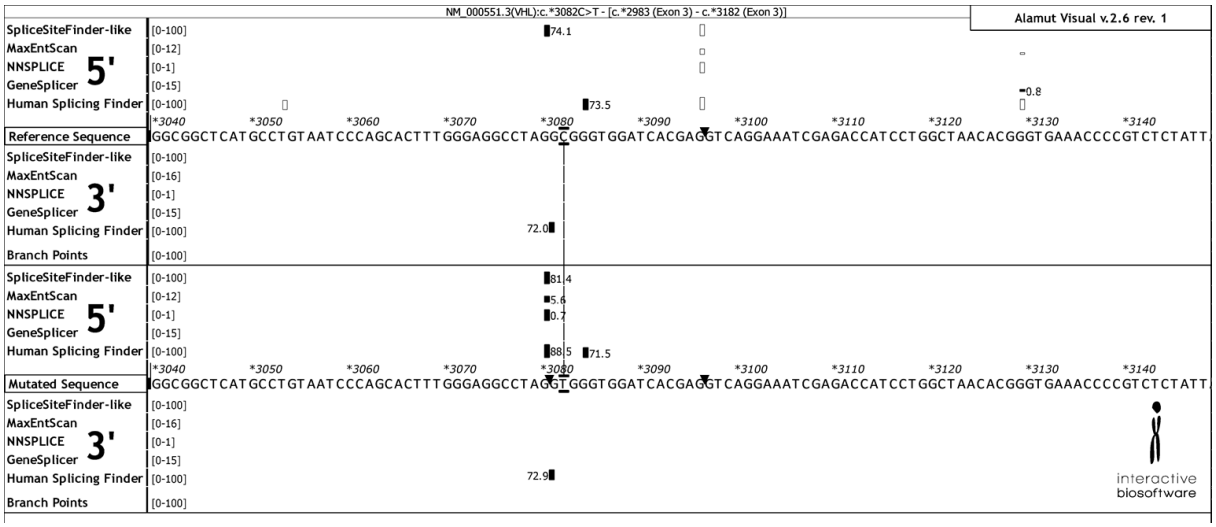
VHL c.*2511A>G



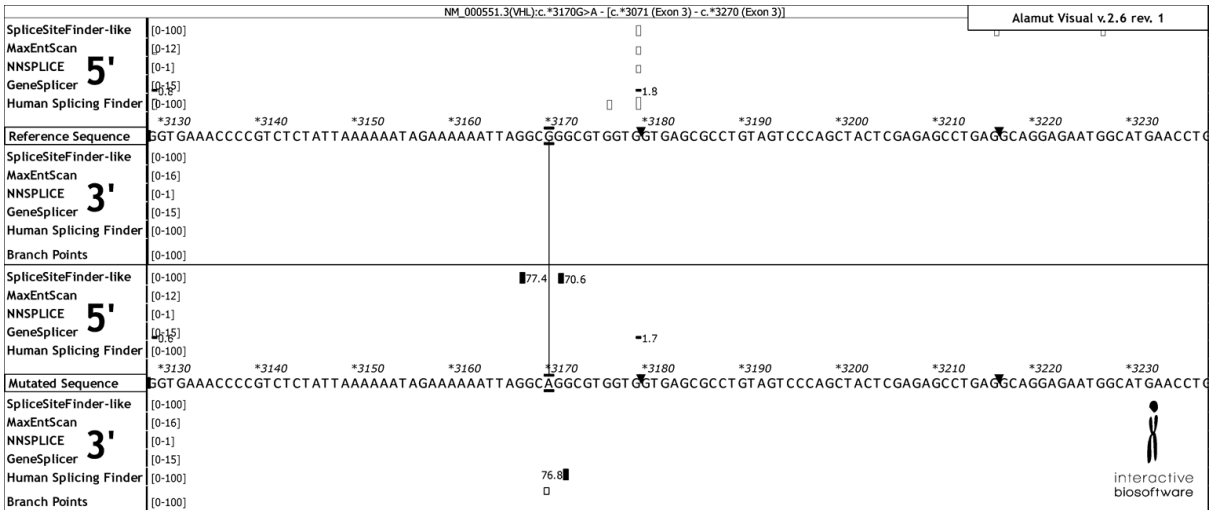
VHL c.*3021T>C



VHL c.*3082C>T



VHL c.*3170G>A



VHL c.*3482dupA

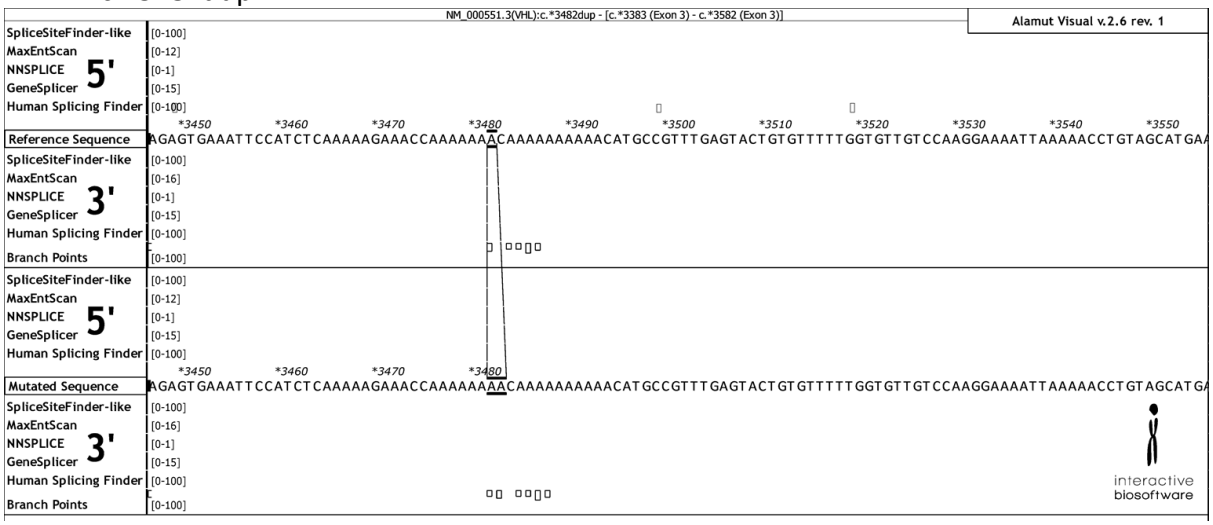


Table 7.5 Calculations for comparison of predicted cryptic splice site scores for sequence variants to average true splice site scores
A summary of these calculations are found in Table 5.31

Gene		SDHD	TMEM127	VHL	VHL	VHL
Variant (coding DNA)		c.[34G>A];[=]	c.[534C>T];[=]	c.[183C>G];[=]	c.[629G>A];[=]	c.[*3082C>T];[=]
Variant (protein)		p.(Gly12Ser)	p.(=)	p.(Pro61Pro)	p.(Arg210Gln)	
Prediction		3' Acceptor created	3' Acceptor created	5' Donor created	3' Acceptor created	5' Donor created
Variant scores	SSF	0.0	81.3	0.0	73.7	81.4
	MES	1.9	4.0	1.6	3.7	5.6
	NNS	0.0	0.0	0.0	0.0	0.7
	GS	0.0	2.9	2.2	0.0	0.0
	HSF	85.3	85.1	71.0	80.3	88.5
Average scores of real splice sites (5' or 3' as appropriate)	SSF	84.3	77.7	83.5	76.5	83.5
	MES	7.5	7.8	10.7	9.2	10.7
	NNS	1.0	0.7	0.9	0.9	0.9
	GS	7.2	8.7	9.1	6.7	9.1
	HSF	87.8	84.8	87.7	82.2	87.7
Average scores less two standard deviations of real splice sites (5' or 3' as appropriate)	SSF	70.5	60.0	63.3	67.5	63.3
	MES	2.3	2.4	8.6	5.4	8.6
	NNS	0.9	-0.5	0.6	0.6	0.6
	GS	3.8	4.2	-10.0	5.1	-10.0
	HSF	79.6	70.8	67.2	80.4	67.2
Variant score greater than average less 2SD of real sites	SSF	NO	YES	NO	YES	YES
	MES	NO	YES	NO	NO	NO
	NNS	NO	YES	NO	NO	YES

Gene		SDHD	TMEM127	VHL	VHL	VHL
Variant (coding DNA)		c.[34G>A];[=]	c.[534C>T];[=]	c.[183C>G];[=]	c.[629G>A];[=]	c.[*3082C>T];[=]
Variant (protein)		p.(Gly12Ser)	p.(=)	p.(Pro61Pro)	p.(Arg210Gln)	
Prediction		3' Acceptor created	3' Acceptor created	5' Donor created	3' Acceptor created	5' Donor created
	GS	NO	NO	YES	NO	YES
	HSF	YES	YES	YES	NO	YES
Number of algorithms for which variant score is equal to or greater than the average scores less 2SDs of the real splice site of that type for that gene		1 of 5	4 of 5	2 of 5	1 of 5	4 of 5

Chapter 8 References

Reference List

(1) White D, Rabago-Smith M. Genotype-phenotype associations and human eye color. J Hum Genet 2011 Jan;56(1):5-7.

(2) Rahman N. Realizing the promise of cancer predisposition genes. Nature 2014 Jan 16;505(7483):302-8.

(3) Durmaz AA, Karaca E, Demkow U, Toruner G, Schoumans J, Cogulu O. Evolution of genetic techniques: past, present, and beyond. Biomed Res Int 2015;2015:461524. doi: 10.1155/2015/461524. Epub@2015 Mar 22.:461524.

(4) Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A 1977 Feb;74(2):560-4.

(5) Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977 Dec;74(12):5463-7.

(6) Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? J Med Genet 2011 Sep;48(9):580-9.

(7) Inoue T, Orgel LE. A nonenzymatic RNA polymerase model. Science 1983 Feb 18;219(4586):859-62.

(8) Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clin Chem 2009 Apr;55(4):641-58.

(9) Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. Am J Hum Genet 2009 Aug;85(2):142-54.

(10) Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet 2010 Oct 15;19(R2):R227-R240.

(11) Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010 Jan;11(1):31-46.

(12) Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012;2012:251364. doi: 10.1155/2012/251364. Epub@2012 Jul 5.:251364.

(13) Binladen J, Gilbert MT, Bollback JP, Panitz F, Bendixen C, Nielsen R, et al. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One 2007 Feb 14;2(2):e197.

(14) Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, et al. Identification of genetic variants using bar-coded multiplexed sequencing. Nat Methods 2008 Oct;5(10):887-93.

(15) Fluidigm. Access Array 48.48 for Roche 454 Sequencing. 2011.
Ref Type: Pamphlet

(16) IIIumina. Nextera XT DNA Sample Preparation Guide. 2012.
Ref Type: Pamphlet

(17) Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 2010;11(12):R119-11.

(18) ten Bosch JR, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. J Mol Diagn 2008 Nov;10(6):484-92.

(19) Shearer AE, DeLuca AP, Hildebrand MS, Taylor KR, Gurrola J, Scherer S, et al. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. Proc Natl Acad Sci U S A 2010 Dec 7;107(49):21104-9.

(20) Read A, Donnai D. Is Cancer Genetic? New Clinical Genetics. Scion Publishing Ltd; 2007. p. 293-328.

(21) de Groot JW, Links TP, Plukker JT, Lips CJ, Hofstra RM. RET as a diagnostic and therapeutic target in sporadic and hereditary endocrine tumors. Endocr Rev 2006 Aug;27(5):535-60.

(22) Storlazzi CT, Lonoce A, Guastadisegni MC, Trombetta D, D'Addabbo P, Daniele G, et al. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. Genome Res 2010 Sep;20(9):1198-206.

(23) Beltran H. The N-myc Oncogene: Maximizing its Targets, Regulation, and Therapeutic Potential. Mol Cancer Res 2014 Jun;12(6):815-22.

(24) Druker BJ. Translation of the Philadelphia chromosome into therapy for CML. Blood 2008 Dec 15;112(13):4808-17.

(25) Sewastianik T, Prochorec-Sobieszek M, Chapuy B, Juszczynski P. MYC deregulation in lymphoid tumors: molecular mechanisms, clinical consequences and therapeutic implications. Biochim Biophys Acta 2014 Dec;1846(2):457-67.

(26) Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971 Apr;68(4):820-3.

(27) Payne SR, Kemp CJ. Tumor suppressor genetics. Carcinogenesis 2005 Dec;26(12):2031-45.

(28) Burnichon N, Cascon A, Schiavi F, Morales NP, Comino-Mendez I, Abermil N, et al. MAX mutations cause hereditary and sporadic pheochromocytoma and paraganglioma. Clin Cancer Res 2012 May 15;18(10):2828-37.

(29) Rahman N, Scott RH. Cancer genes associated with phenotypes in monoallelic and biallelic mutation carriers: new lessons from old players. Hum Mol Genet 2007 Apr 15;16 Spec No 1:R60-6.:R60-R66.

(30) Astuti D, Latif F, Dallol A, Dahia PL, Douglas F, George E, et al. Gene mutations in the succinate dehydrogenase subunit SDHB cause susceptibility to familial pheochromocytoma and to familial paraganglioma. Am J Hum Genet 2001 Jul;69(1):49-54.

(31) Davit-Spraul A, Gonzales E, Baussan C, Jacquemin E. Progressive familial intrahepatic cholestasis. Orphanet J Rare Dis 2009 Jan 8;4:1. doi: 10.1186/1750-1172-4-1.:1-4.

(32) Dimaras H, Kimani K, Dimba EA, Gronsdahl P, White A, Chan HS, et al. Retinoblastoma. Lancet 2012 Apr 14;379(9824):1436-46.

(33) Jafri M, Maher ER. The genetics of phaeochromocytoma: using clinical features to guide genetic testing. Eur J Endocrinol 2012 Feb;166(2):151-8.

(34) Rushlow D, Piovesan B, Zhang K, Prigoda-Lee NL, Marchong MN, Clark RD, et al. Detection of mosaic RB1 mutations in families with retinoblastoma. Hum Mutat 2009 May;30(5):842-51.

(35) Cancer risks in BRCA2 mutation carriers. J Natl Cancer Inst 1999 Aug 4;91(15):1310-6.

(36) Thompson D, Easton D. Variation in cancer risks, by mutation position, in BRCA2 mutation carriers. Am J Hum Genet 2001 Feb;68(2):410-9.

(37) Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 2005 Apr 14;434(7035):917-21.

(38) Metcalfe K, Gershman S, Ghadirian P, Lynch HT, Snyder C, Tung N, et al. Contralateral mastectomy and survival after breast cancer in carriers of BRCA1 and BRCA2 mutations: retrospective analysis. BMJ 2014 Feb 11;348:g226. doi: 10.1136/bmj.g226.:g226.

(39) van der Post RS, Vogelaar IP, Manders P, van der Kolk LE, Cats A, van Hest LP, et al. Accuracy of Hereditary Diffuse Gastric Cancer Testing Criteria and Outcomes in Patients with a Germline Mutation in CDH1. Gastroenterology 2015 Jun 10;(15):10.

(40) Chen H, Sippel RS, O'Dorisio MS, Vinik AI, Lloyd RV, Pacak K. The North American Neuroendocrine Tumor Society consensus guideline for the diagnosis and

management of neuroendocrine tumors: pheochromocytoma, paraganglioma, and medullary thyroid cancer. Pancreas 2010 Aug;39(6):775-83.

(41) Smith A, Moran A, Boyd MC, Bulman M, Shenton A, Smith L, et al. Phenocopies in BRCA1 and BRCA2 families: evidence for modifier genes and implications for screening. J Med Genet 2007 Jan;44(1):10-5.

(42) Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000;132:365-86.:365-86.

(43) Martins R, Bugalho MJ. Paragangliomas/Pheochromocytomas: clinically oriented genetic testing. Int J Endocrinol 2014;2014:794187. doi: 10.1155/2014/794187. Epub@2014 May 12.:794187.

(44) DeLellis R, Lloyd R, Heitz P, Eng C. Pathology and Genetics of Tumours of Endocrine Organs (IARC WHO Classification of Tumours). 2004.

(45) Lenders JW, Eisenhofer G, Mannelli M, Pacak K. Phaeochromocytoma. Lancet 2005 Aug;366(9486):665-75.

(46) Harari A, Inabnet WB, III. Malignant pheochromocytoma: a review. Am J Surg 2011 May;201(5):700-8.

(47) King KS, Prodanov T, Kantorovich V, Fojo T, Hewitt JK, Zacharin M, et al. Metastatic pheochromocytoma/paraganglioma related to primary tumor development in childhood or adolescence: significant link to SDHB mutations. J Clin Oncol 2011 Nov 1;29(31):4137-42.

(48) Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. BMJ 2014 Nov 10;349:g4797. doi: 10.1136/bmj.g4797.:g4797.

(49) Maher ER. Genetics of familial renal cancers. Nephron Exp Nephrol 2011;118(1):e21-e26.

(50) Morrison PJ, Donnelly DE, Atkinson AB, Maxwell AP. Advances in the genetics of familial renal cancer. Oncologist 2010;15(6):532-8.

(51) Hagenkord JM, Gatalica Z, Jonasch E, Monzon FA. Clinical genomics of renal epithelial tumors. Cancer Genet 2011 Jun;204(6):285-97.

(52) Gimenez-Roqueplo AP, Lehnert H, Mannelli M, Neumann H, Opocher G, Maher ER, et al. Phaeochromocytoma, new genes and screening strategies. Clin Endocrinol (Oxf) 2006 Dec;65(6):699-705.

(53) Karasek D, Frysak Z, Pacak K. Genetic testing for pheochromocytoma. Curr Hypertens Rep 2010 Dec;12(6):456-64.

(54) Erlic Z, Rybicki L, Peczkowska M, Golcher H, Kann PH, Brauckhoff M, et al. Clinical predictors and algorithm for the genetic diagnosis of pheochromocytoma patients. Clin Cancer Res 2009 Oct 15;15(20):6378-85.

(55) Neumann HP, Erlic Z, Boedeker CC, Rybicki LA, Robledo M, Hermsen M, et al. Clinical predictors for germline mutations in head and neck paraganglioma patients: cost reduction strategy in genetic diagnostic process as fall-out. Cancer Res 2009 Apr 15;69(8):3650-6.

(56) Jafri M, Whitworth J, Rattenberry E, Vialard L, Kilby G, Kumar AV, et al. Evaluation of SDHB, SDHD and VHL gene susceptibility testing in the assessment of individuals with non-syndromic phaeochromocytoma, paraganglioma and head and neck paraganglioma. Clin Endocrinol (Oxf) 2013 Jun;78(6):898-906.

(57) Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev 2011 Nov;32(4):177-95.

(58) Morgan JE, Carr IM, Sheridan E, Chu CE, Hayward B, Camm N, et al. Genetic diagnosis of familial breast cancer using clonal sequencing. Hum Mutat 2010 Apr;31(4):484-91.

(59) O'Sullivan J, Mullaney BG, Bhaskar SS, Dickerson JE, Hall G, O'Grady A, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. J Med Genet 2012 May;49(5):322-6.

(60) Verine J, Pluvinage A, Bousquet G, Lehmann-Che J, de BC, Soufir N, et al. Hereditary renal cancer syndromes: an update of a systematic review. Eur Urol 2010 Nov;58(5):701-10.

(61) Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005 Sep 15;437(7057):376-80.

(62) Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. Nucleic Acids Res 2012 Aug;40(15):e115.

(63) Bell JA, Bodmer D, Sistermans E, Ramsden SC. Practice guidelines for the Interpretation and Reporting of Unclassified Variants (UVs) in Clinical Molecular Genetics. 22-10-2007.
Ref Type: Online Source

(64) Astuti D, Ricketts CJ, Chowdhury R, McDonough MA, Gentle D, Kirby G, et al. Mutation analysis of HIF prolyl hydroxylases (PHD/EGLN) in individuals with features of phaeochromocytoma and renal cell carcinoma susceptibility. Endocr Relat Cancer 2010 Dec 21;18(1):73-83.

(65) Ladroue C, Carcenac R, Leporrier M, Gad S, Le HC, Galateau-Salle F, et al. PHD2 mutation and congenital erythrocytosis with paraganglioma. N Engl J Med 2008 Dec 18;359(25):2685-92.

(66) Schlisio S, Kenchappa RS, Vredeveld LC, George RE, Stewart R, Greulich H, et al. The kinesin KIF1Bbeta acts downstream from EglN3 to induce apoptosis and is a potential 1p36 tumor suppressor. Genes Dev 2008 Apr 1;22(7):884-93.

(67) Zhao C, Takita J, Tanaka Y, Setou M, Nakagawa T, Takeda S, et al. Charcot-Marie-Tooth disease type 2A caused by mutation in a microtubule motor KIF1Bbeta. Cell 2001 Jun 1;105(5):587-97.

(68) Comino-Mendez I, Gracia-Aznarez FJ, Schiavi F, Landa I, Leandro-Garcia LJ, Leton R, et al. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. Nat Genet 2011 Jun;%19;43(7):663-7.

(69) Fishbein L, Nathanson KL. Pheochromocytoma and paraganglioma: understanding the complexities of the genetic background. Cancer Genet 2012 Jan;205(1-2):1-11.

(70) Bausch B, Borozdin W, Mautner VF, Hoffmann MM, Boehm D, Robledo M, et al. Germline NF1 mutational spectra and loss-of-heterozygosity analyses in patients with pheochromocytoma and neurofibromatosis type 1. J Clin Endocrinol Metab 2007 Jul;92(7):2784-92.

(71) Phay JE, Shah MH. Targeting RET receptor tyrosine kinase activation in cancer. Clin Cancer Res 2010 Dec 15;16(24):5936-41.

(72) Burnichon N, Briere JJ, Libe R, Vescovo L, Riviere J, Tissier F, et al. SDHA is a tumor suppressor gene causing paraganglioma. Hum Mol Genet 2010 Aug 1;19(15):3011-20.

(73) Korpershoek E, Favier J, Gaal J, Burnichon N, van GB, Oudijk L, et al. SDHA immunohistochemistry detects germline SDHA gene mutations in apparently sporadic paragangliomas and pheochromocytomas. J Clin Endocrinol Metab 2011 Sep;96(9):E1472-E1476.

(74) Baysal BE, Lawrence EC, Ferrell RE. Sequence variation in human succinate dehydrogenase genes: evidence for long-term balancing selection on SDHA. BMC Biol 2007 Mar 21;5:12.:12.

(75) Ricketts CJ, Forman JR, Rattenberry E, Bradshaw N, Lalloo F, Izatt L, et al. Tumor risks and genotype-phenotype-proteotype analysis in 358 patients with germline mutations in SDHB and SDHD. Hum Mutat 2010 Jan;31(1):41-51.

(76) Niemann S, Muller U. Mutations in SDHC cause autosomal dominant paraganglioma, type 3. Nat Genet 2000 Nov;26(3):268-70.

(77) Schiavi F, Boedeker CC, Bausch B, Peczkowska M, Gomez CF, Strassburg T, et al. Predictors and prevalence of paraganglioma syndrome associated with mutations of the SDHC gene. JAMA 2005 Oct 26;294(16):2057-63.

(78) Vandy FC, Sisk G, Berguer R. Synchronous carotid body and thoracic paraganglioma associated with a germline SDHC mutation. J Vasc Surg 2011 Mar;53(3):805-7.

(79) Baysal BE, Ferrell RE, Willett-Brozick JE, Lawrence EC, Myssiorek D, Bosch A, et al. Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. Science 2000 Feb 4;287(5454):848-51.

(80) Pigny P, Vincent A, Cardot BC, Bertrand M, de Montpreville VT, Crepin M, et al. Paraganglioma after maternal transmission of a succinate dehydrogenase gene mutation. J Clin Endocrinol Metab 2008 May;93(5):1609-15.

(81) Bayley JP, Kunst HP, Cascon A, Sampietro ML, Gaal J, Korpershoek E, et al. SDHAF2 mutations in familial and sporadic paraganglioma and phaeochromocytoma. Lancet Oncol 2010 Apr;11(4):366-72.

(82) Kunst HP, Rutten MH, de Monnink JP, Hoefsloot LH, Timmers HJ, Marres HA, et al. SDHAF2 (PGL2-SDH5) and hereditary head and neck paraganglioma. Clin Cancer Res 2011 Jan 15;17(2):247-54.

(83) Yao L, Barontini M, Niederle B, Jech M, Pfragner R, Dahia PL. Mutations of the metabolic genes IDH1, IDH2, and SDHAF2 are not major determinants of the pseudohypoxic phenotype of sporadic pheochromocytomas and paragangliomas. J Clin Endocrinol Metab 2010 Mar;95(3):1469-72.

(84) Yao L, Schiavi F, Cascon A, Qin Y, Inglada-Perez L, King EE, et al. Spectrum and prevalence of FP/TMEM127 gene mutations in pheochromocytomas and paragangliomas. JAMA 2010 Dec 15;304(23):2611-9.

(85) Neumann HP, Sullivan M, Winter A, Malinoc A, Hoffmann MM, Boedeker CC, et al. Germline mutations of the TMEM127 gene in patients with paraganglioma of head and neck and extraadrenal abdominal sites. J Clin Endocrinol Metab 2011 Aug;96(8):E1279-E1282.

(86) Richards FM. Molecular pathology of von HippelLindau disease and the VHL tumour suppressor gene. Expert Rev Mol Med 2001 Mar;2001:1-27.:1-27.

(87) Pacak K, Eisenhofer G, Ahlman H, Bornstein SR, Gimenez-Roqueplo AP, Grossman AB, et al. Pheochromocytoma: recommendations for clinical practice from the First International Symposium. October 2005. Nat Clin Pract Endocrinol Metab 2007 Feb;3(2):92-102.

(88) Lonser RR, Glenn GM, Walther M, Chew EY, Libutti SK, Linehan WM, et al. von Hippel-Lindau disease. Lancet 2003 Jun 14;361(9374):2059-67.

(89) Schmidt L, Duh FM, Chen F, Kishida T, Glenn G, Choyke P, et al. Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. Nat Genet 1997 May;16(1):68-73.

(90) Tomlinson IP, Alam NA, Rowan AJ, Barclay E, Jaeger EE, Kelsell D, et al. Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. Nat Genet 2002 Apr;30(4):406-10.

(91) Baba M, Hong SB, Sharma N, Warren MB, Nickerson ML, Iwamatsu A, et al. Folliculin encoded by the BHD gene interacts with a binding protein, FNIP1, and AMPK, and is involved in AMPK and mTOR signaling. Proc Natl Acad Sci U S A 2006 Oct 17;103(42):15552-7.

(92) Nickerson ML, Warren MB, Toro JR, Matrosova V, Glenn G, Turner ML, et al. Mutations in a novel gene lead to kidney tumors, lung wall defects, and benign tumors of the hair follicle in patients with the Birt-Hogg-Dube syndrome. Cancer Cell 2002 Aug;2(2):157-64.

(93) Identification and characterization of the tuberous sclerosis gene on chromosome 16. Cell 1993 Dec 31;75(7):1305-15.

(94) van SM, de HR, Hermans C, Nellist M, Janssen B, Verhoef S, et al. Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34. Science 1997 Aug 8;277(5327):805-8.

(95) Rozenblatt-Rosen O, Hughes CM, Nannepaga SJ, Shanmugam KS, Copeland TD, Guszczynski T, et al. The parafibromin tumor suppressor protein is part of a human Paf1 complex. Mol Cell Biol 2005 Jan;25(2):612-20.

(96) Carpten JD, Robbins CM, Villablanca A, Forsberg L, Presciuttini S, Bailey-Wilson J, et al. HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. Nat Genet 2002 Dec;32(4):676-80.

(97) Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 2007;8(7):R143.

(98) De Leeneer K, De SJ, Clement L, Baetens M, Lefever S, De KS, et al. Practical tools to implement massive parallel pyrosequencing of PCR products in next generation molecular diagnostics. PLoS One 2011;6(9):e25531.

(99) Hollants S, Redeker EJ, Matthijs G. Microfluidic amplification as a tool for massive parallel sequencing of the familial hypercholesterolemia genes. Clin Chem 2012 Apr;58(4):717-24.

(100) Michils G, Hollants S, Dehaspe L, Van HJ, Bidet Y, Uhrhammer N, et al. Molecular analysis of the breast cancer genes BRCA1 and BRCA2 using amplicon-based massive parallel pyrosequencing. J Mol Diagn 2012 Nov;14(6):623-30.

(101) Astuti D, Douglas F, Lennard TW, Aligianis IA, Woodward ER, Evans DG, et al. Germline SDHD mutation in familial phaeochromocytoma. Lancet 2001 Apr 14;357(9263):1181-2.

(102) Bonnal RJ, Severgnini M, Castaldi A, Bordoni R, Iacono M, Trimarco A, et al. Reliable resequencing of the human dystrophin locus by universal long polymerase chain reaction and massive pyrosequencing. Anal Biochem 2010 Nov 15;406(2):176-84.

(103) De Leeneer K, Hellemans J, De SJ, Baetens M, Poppe B, Van CW, et al. Massive parallel amplicon sequencing of the breast cancer genes BRCA1 and BRCA2: opportunities, challenges, and limitations. Hum Mutat 2011 Mar;32(3):335-44.

(104) Jafri M, Whitworth J, Rattenberry E, Bradley L, Kilby G, Kumar AV, et al. Evaluation of SDHB, SDHD and VHL gene susceptibility testing in the assessment of individuals with non-syndromic phaeochromocytoma and paraganglioma and head and neck paraganglioma (HNPGL). Clin Endocrinol (Oxf) 2012 Oct 16;10.

(105) Henderson A, Douglas F, Perros P, Morgan C, Maher ER. SDHB-associated renal oncocytoma suggests a broadening of the renal phenotype in hereditary paragangliomatosis. Fam Cancer 2009;8(3):257-60.

(106) Gossage L, Eisen T, Maher ER. VHL, the story of a tumour suppressor gene. Nat Rev Cancer 2015 Jan;15(1):55-64.

(107) Maher ER, Iselius L, Yates JR, Littler M, Benjamin C, Harris R, et al. Von Hippel-Lindau disease: a genetic study. J Med Genet 1991 Jul;28(7):443-7.

(108) Von Hippel E. Ueber eine sehr seltene Erkrankung der Netzhaut. Albrecht von Graefes Arch Ophthal 1904;59:83-106.

(109) Lindau A. Zur Frage der Angiomatosis Retinae und Ihrer Hirncomplikation. Acta Opthalmol Scand 1927;4:193-226.

(110) MELMON KL, ROSEN SW. LINDAU'S DISEASE. REVIEW OF THE LITERATURE AND STUDY OF A LARGE KINDRED. Am J Med 1964 Apr;36:595-617.:595-617.

(111) Latif F, Tory K, Gnarra J, Yao M, Duh FM, Orcutt ML, et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. Science 1993 May 28;260(5112):1317-20.

(112) Tory K, Brauch H, Linehan M, Barba D, Oldfield E, Filling-Katz M, et al. Specific genetic change in tumors associated with von Hippel-Lindau disease. J Natl Cancer Inst 1989 Jul 19;81(14):1097-101.

(113) Crossey PA, Foster K, Richards FM, Phipps ME, Latif F, Tory K, et al. Molecular genetic investigations of the mechanism of tumourigenesis in von Hippel-Lindau disease: analysis of allele loss in VHL tumours. Hum Genet 1994 Jan;93(1):53-8.

(114) Zbar B, Brauch H, Talmadge C, Linehan M. Loss of alleles of loci on the short arm of chromosome 3 in renal cell carcinoma. Nature 1987 Jun 25;327(6124):721-4.

(115) Banks RE, Tirukonda P, Taylor C, Hornigold N, Astuti D, Cohen D, et al. Genetic and epigenetic analysis of von Hippel-Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. Cancer Res 2006 Feb 15;66(4):2000-11.

(116) Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A 1994 Oct 11;91(21):9700-4.

(117) Iliopoulos O, Kibel A, Gray S, Kaelin WG, Jr. Tumour suppression by the human von Hippel-Lindau gene product. Nat Med 1995 Aug;1(8):822-6.

(118) Pause A, Lee S, Lonergan KM, Klausner RD. The von Hippel-Lindau tumor suppressor gene is required for cell cycle exit upon serum withdrawal. Proc Natl Acad Sci U S A 1998 Feb 3;95(3):993-8.

(119) Neumann HP, Wiestler OD. Clustering of features of von Hippel-Lindau syndrome: evidence for a complex genetic locus. Lancet 1991 May 4;337(8749):1052-4.

(120) Glenn GM, Daniel LN, Choyke P, Linehan WM, Oldfield E, Gorin MB, et al. Von Hippel-Lindau (VHL) disease: distinct phenotypes suggest more than one mutant allele at the VHL locus. Hum Genet 1991 Jun;87(2):207-10.

(121) Hoffman MA, Ohh M, Yang H, Klco JM, Ivan M, Kaelin WG, Jr. von Hippel-Lindau protein mutants linked to type 2C VHL disease preserve the ability to downregulate HIF. Hum Mol Genet 2001 May 1;10(10):1019-27.

(122) Ang SO, Chen H, Hirota K, Gordeuk VR, Jelinek J, Guan Y, et al. Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. Nat Genet 2002 Dec;32(4):614-21.

(123) Richard S, Gardie B, Couve S, Gad S. Von Hippel-Lindau: how a rare disease illuminates cancer biology. Semin Cancer Biol 2013 Feb;23(1):26-37.

(124) Tomasic NL, Piterkova L, Huff C, Bilic E, Yoon D, Miasnikova GY, et al. The phenotype of polycythemia due to Croatian homozygous VHL (571C>G:H191D) mutation is different from that of Chuvash polycythemia (VHL 598C>T:R200W). Haematologica 2013 Apr;98(4):560-7.

(125) Lanikova L, Lorenzo F, Yang C, Vankayalapati H, Drachtman R, Divoky V, et al. Novel homozygous VHL mutation in exon 2 is associated with congenital polycythemia but not with cancer. Blood 2013 May 9;121(19):3918-24.

(126) Bento C, Almeida H, Maia TM, Relvas L, Oliveira AC, Rossi C, et al. Molecular study of congenital erythrocytosis in 70 unrelated patients revealed a potential causal

mutation in less than half of the cases (Where is/are the missing gene(s)?). Eur J Haematol 2013 Oct;91(4):361-8.

(127) Pastore YD, Jelinek J, Ang S, Guan Y, Liu E, Jedlickova K, et al. Mutations in the VHL gene in sporadic apparently congenital polycythemia. Blood 2003 Feb 15;101(4):1591-5.

(128) Bento C, Percy MJ, Gardie B, Maia TM, van WR, Perrotta S, et al. Genetic basis of congenital erythrocytosis: mutation update and online databases. Hum Mutat 2014 Jan;35(1):15-26.

(129) Capodimonti S, Teofili L, Martini M, Cenci T, Iachininoto MG, Nuzzolo ER, et al. Von hippel-lindau disease and erythrocytosis. J Clin Oncol 2012 May 1;30(13):e137-e139.

(130) Hickey MM, Lam JC, Bezman NA, Rathmell WK, Simon MC. von Hippel-Lindau mutation in mice recapitulates Chuvash polycythemia via hypoxia-inducible factor-2alpha signaling and splenic erythropoiesis. J Clin Invest 2007 Dec;117(12):3879-89.

(131) Gordeuk VR, Sergueeva AI, Miasnikova GY, Okhotin D, Voloshin Y, Choyke PL, 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 May 15;103(10):3924-32.

(132) Miasnikova GY, Sergueeva AI, Nouraie M, Niu X, Okhotin DJ, Polyakova LA, et al. The heterozygote advantage of the Chuvash polycythemia VHLR200W mutation may be protection against anemia. Haematologica 2011 Sep;96(9):1371-4.

(133) Da Silva JL, Lacombe C, Bruneval P, Casadevall N, Leporrier M, Camilleri JP, et al. Tumor cells are the site of erythropoietin synthesis in human renal cancers associated with polycythemia. Blood 1990 Feb 1;75(3):577-82.

(134) Krieg M, Marti HH, Plate KH. Coexpression of erythropoietin and vascular endothelial growth factor in nervous system tumors associated with von Hippel-Lindau tumor suppressor gene loss of function. Blood 1998 Nov 1;92(9):3388-93.

(135) Gnarra JR, Tory K, Weng Y, Schmidt L, Wei MH, Li H, et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. Nat Genet 1994 May;7(1):85-90.

(136) Richards FM, Schofield PN, Fleming S, Maher ER. Expression of the von Hippel-Lindau disease tumour suppressor gene during human embryogenesis. Hum Mol Genet 1996 May;5(5):639-44.

(137) Whaley JM, Naglich J, Gelbert L, Hsia YE, Lamiell JM, Green JS, et al. Germ-line mutations in the von Hippel-Lindau tumor-suppressor gene are similar to somatic von Hippel-Lindau aberrations in sporadic renal cell carcinoma. Am J Hum Genet 1994 Dec;55(6):1092-102.

(138) Richards FM, Crossey PA, Phipps ME, Foster K, Latif F, Evans G, et al. Detailed mapping of germline deletions of the von Hippel-Lindau disease tumour suppressor gene. Hum Mol Genet 1994 Apr;3(4):595-8.

(139) Franke G, Bausch B, Hoffmann MM, Cybulla M, Wilhelm C, Kohlhase J, et al. Alu-Alu recombination underlies the vast majority of large VHL germline deletions: Molecular characterization and genotype-phenotype correlations in VHL patients. Hum Mutat 2009 May;30(5):776-86.

(140) Blankenship C, Naglich JG, Whaley JM, Seizinger B, Kley N. Alternate choice of initiation codon produces a biologically active product of the von Hippel Lindau gene with tumor suppressor activity. Oncogene 1999 Feb 25;18(8):1529-35.

(141) Iliopoulos O, Ohh M, Kaelin WG, Jr. 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 Sep 29;95(20):11661-6.

(142) Schoenfeld A, Davidowitz EJ, Burk RD. 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 Jul 21;95(15):8817-22.

(143) Hergovich A, Lisztwan J, Barry R, Ballschmieter P, Krek W. Regulation of microtubule stability by the von Hippel-Lindau tumour suppressor protein pVHL. Nat Cell Biol 2003 Jan;5(1):64-70.

(144) Nordstrom-O'Brien M, van der Luijt RB, van RE, van den Ouweland AM, Majoor-Krakauer DF, Lolkema MP, et al. Genetic analysis of von Hippel-Lindau disease. Hum Mutat 2010 May;31(5):521-37.

(145) Robinson CM, Ohh M. The multifaceted von Hippel-Lindau tumour suppressor protein. FEBS Lett 2014 Aug;588(16):2704-11.

(146) Duan DR, Pause A, Burgess WH, Aso T, Chen DY, Garrett KP, et al. Inhibition of transcription elongation by the VHL tumor suppressor protein. Science 1995 Sep 8;269(5229):1402-6.

(147) Kibel A, Iliopoulos O, DeCaprio JA, Kaelin WG, Jr. Binding of the von Hippel-Lindau tumor suppressor protein to Elongin B and C. Science 1995 Sep 8;269(5229):1444-6.

(148) Kishida T, Stackhouse TM, Chen F, Lerman MI, Zbar B. Cellular proteins that bind the von Hippel-Lindau disease gene product: mapping of binding domains and the effect of missense mutations. Cancer Res 1995 Oct 15;55(20):4544-8.

(149) Stebbins CE, Kaelin WG, Jr., Pavletich NP. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. Science 1999 Apr 16;284(5413):455-61.

(150) Lonergan KM, Iliopoulos O, Ohh M, Kamura T, Conaway RC, Conaway JW, et al. Regulation of hypoxia-inducible mRNAs by the von Hippel-Lindau tumor suppressor protein requires binding to complexes containing elongins B/C and Cul2. Mol Cell Biol 1998 Feb;18(2):732-41.

(151) Kamura T, Koepp DM, Conrad MN, Skowyra D, Moreland RJ, Iliopoulos O, et al. Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. Science 1999 Apr 23;284(5414):657-61.

(152) Pause A, Lee S, Worrell RA, Chen DY, Burgess WH, Linehan WM, et al. The von Hippel-Lindau tumor-suppressor gene product forms a stable complex with human CUL-2, a member of the Cdc53 family of proteins. Proc Natl Acad Sci U S A 1997 Mar 18;94(6):2156-61.

(153) Schoenfeld AR, Davidowitz EJ, Burk RD. Elongin BC complex prevents degradation of von Hippel-Lindau tumor suppressor gene products. Proc Natl Acad Sci U S A 2000 Jul 18;97(15):8507-12.

(154) Hon WC, Wilson MI, Harlos K, Claridge TD, Schofield CJ, Pugh CW, et al. Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. Nature 2002 Jun 27;417(6892):975-8.

(155) Min JH, Yang H, Ivan M, Gertler F, Kaelin WG, Jr., Pavletich NP. Structure of an HIF-1alpha -pVHL complex: hydroxyproline recognition in signaling. Science 2002 Jun 7;296(5574):1886-9.

(156) Knauth K, Bex C, Jemth P, Buchberger A. Renal cell carcinoma risk in type 2 von Hippel-Lindau disease correlates with defects in pVHL stability and HIF-1alpha interactions. Oncogene 2006 Jan;25(3):370-7.

(157) Clifford SC, Cockman ME, Smallwood AC, Mole DR, Woodward ER, Maxwell PH, 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 May 1;10(10):1029-38.

(158) Li L, Zhang L, Zhang X, Yan Q, Minamishima YA, Olumi AF, et al. Hypoxia-inducible factor linked to differential kidney cancer risk seen with type 2A and type 2B VHL mutations. Mol Cell Biol 2007 Aug;27(15):5381-92.

(159) Ivan M, Kondo K, Yang H, Kim W, Valiando J, Ohh M, et al. HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. Science 2001 Apr;292(5516):464-8.

(160) Bruick RK, McKnight SL. A conserved family of prolyl-4-hydroxylases that modify HIF. Science 2001 Nov 9;294(5545):1337-40.

(161) Jaakkola P, Mole DR, Tian YM, Wilson MI, Gielbert J, Gaskell SJ, et al. Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. Science 2001 Apr;292(5516):468-72.

(162) Yu F, White SB, Zhao Q, Lee FS. HIF-1alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation. Proc Natl Acad Sci U S A 2001 Aug 14;98(17):9630-5.

(163) Semenza GL. Hypoxia-inducible factors: mediators of cancer progression and targets for cancer therapy. Trends Pharmacol Sci 2012 Apr;33(4):207-14.

(164) Xia X, Lemieux ME, Li W, Carroll JS, Brown M, Liu XS, et al. Integrative analysis of HIF binding and transactivation reveals its role in maintaining histone methylation homeostasis. Proc Natl Acad Sci U S A 2009 Mar 17;106(11):4260-5.

(165) Schodel J, Oikonomopoulos S, Ragoussis J, Pugh CW, Ratcliffe PJ, Mole DR. High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. Blood 2011 Jun 9;117(23):e207-e217.

(166) Semenza GL. Oxygen sensing, homeostasis, and disease. N Engl J Med 2011 Aug 11;365(6):537-47.

(167) Crosby ME, Devlin CM, Glazer PM, Calin GA, Ivan M. Emerging roles of microRNAs in the molecular responses to hypoxia. Curr Pharm Des 2009;15(33):3861-6.

(168) Wu MZ, Tsai YP, Yang MH, Huang CH, Chang SY, Chang CC, et al. Interplay between HDAC3 and WDR5 is essential for hypoxia-induced epithelial-mesenchymal transition. Mol Cell 2011 Sep 2;43(5):811-22.

(169) Beyer S, Kristensen MM, Jensen KS, Johansen JV, Staller P. The histone demethylases JMJD1A and JMJD2B are transcriptional targets of hypoxia-inducible factor HIF. J Biol Chem 2008 Dec 26;283(52):36542-52.

(170) Pollard PJ, Loenarz C, Mole DR, McDonough MA, Gleadle JM, Schofield CJ, et al. Regulation of Jumonji-domain-containing histone demethylases by hypoxia-inducible factor (HIF)-1alpha. Biochem J 2008 Dec 15;416(3):387-94.

(171) Wiesener MS, Jurgensen JS, Rosenberger C, Scholze CK, Horstrup JH, Warnecke C, et al. Widespread hypoxia-inducible expression of HIF-2alpha in distinct cell populations of different organs. FASEB J 2003 Feb;17(2):271-3.

(172) Keith B, Johnson RS, Simon MC. HIF1alpha and HIF2alpha: sibling rivalry in hypoxic tumour growth and progression. Nat Rev Cancer 2011 Dec 15;12(1):9-22.

(173) Hu CJ, Wang LY, Chodosh LA, Keith B, Simon MC. Differential roles of hypoxia-inducible factor 1alpha (HIF-1alpha) and HIF-2alpha in hypoxic gene regulation. Mol Cell Biol 2003 Dec;23(24):9361-74.

(174) Raval RR, Lau KW, Tran MG, Sowter HM, Mandriota SJ, Li JL, et al. Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma. Mol Cell Biol 2005 Jul;25(13):5675-86.

(175) Carroll VA, Ashcroft M. Role of hypoxia-inducible factor (HIF)-1alpha versus HIF-2alpha in the regulation of HIF target genes in response to hypoxia, insulin-like growth factor-I, or loss of von Hippel-Lindau function: implications for targeting the HIF pathway. Cancer Res 2006 Jun 15;66(12):6264-70.

(176) Singla V, Reiter JF. The primary cilium as the cell's antenna: signaling at a sensory organelle. Science 2006 Aug 4;313(5787):629-33.

(177) Lutz MS, Burk RD. Primary cilium formation requires von hippel-lindau gene function in renal-derived cells. Cancer Res 2006 Jul 15;66(14):6903-7.

(178) Ohh M, Yauch RL, Lonergan KM, Whaley JM, Stemmer-Rachamimov AO, Louis DN, et al. The von Hippel-Lindau tumor suppressor protein is required for proper assembly of an extracellular fibronectin matrix. Mol Cell 1998 Jun;1(7):959-68.

(179) Li M, Kim WY. Two sides to every story: the HIF-dependent and HIF-independent functions of pVHL. J Cell Mol Med 2011 Feb;15(2):187-95.

(180) Yang H, Minamishima YA, Yan Q, Schlisio S, Ebert BL, Zhang X, et al. pVHL acts as an adaptor to promote the inhibitory phosphorylation of the NF-kappaB agonist Card9 by CK2. Mol Cell 2007 Oct 12;28(1):15-27.

(181) Roe JS, Kim H, Lee SM, Kim ST, Cho EJ, Youn HD. p53 stabilization and transactivation by a von Hippel-Lindau protein. Mol Cell 2006 May 5;22(3):395-405.

(182) Young AP, Schlisio S, Minamishima YA, Zhang Q, Li L, Grisanzio C, et al. VHL loss actuates a HIF-independent senescence programme mediated by Rb and p400. Nat Cell Biol 2008 Mar;10(3):361-9.

(183) Kuznetsova AV, Meller J, Schnell PO, Nash JA, Ignacak ML, Sanchez Y, 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 Mar 4;100(5):2706-11.

(184) Beroud C, Joly D, Gallou C, Staroz F, Orfanelli MT, Junien C. Software and database for the analysis of mutations in the VHL gene. Nucleic Acids Res 1998 Jan 1;26(1):256-8.

(185) Cascon A, Escobar B, Montero-Conde C, Rodriguez-Antona C, Ruiz-Llorente S, Osorio A, et al. Loss of the actin regulator HSPC300 results in clear cell renal cell carcinoma protection in Von Hippel-Lindau patients. Hum Mutat 2007 Jun;28(6):613-21.

(186) Maranchie JK, Afonso A, Albert PS, Kalyandrug S, Phillips JL, Zhou S, et al. Solid renal tumor severity in von Hippel Lindau disease is related to germline deletion length and location. Hum Mutat 2004 Jan;23(1):40-6.

(187) McNeill A, Rattenberry E, Barber R, Killick P, Macdonald F, Maher ER. Genotype-phenotype correlations in VHL exon deletions. Am J Med Genet A 2009 Oct;149A(10):2147-51.

(188) Escobar B, de CG, Fernandez-Miranda G, Cascon A, Bravo-Cordero JJ, Montoya MC, et al. Brick1 is an essential regulator of actin cytoskeleton required for embryonic development and cell transformation. Cancer Res 2010 Nov 15;70(22):9349-59.

(189) Ong KR, Woodward ER, Killick P, Lim C, Macdonald F, Maher ER. Genotype-phenotype correlations in von Hippel-Lindau disease. Hum Mutat 2007 Feb;28(2):143-9.

(190) Zatyka M, da Silva NF, Clifford SC, Morris MR, Wiesener MS, Eckardt KU, et al. Identification of cyclin D1 and other novel targets for the von Hippel-Lindau tumor suppressor gene by expression array analysis and investigation of cyclin D1 genotype as a modifier in von Hippel-Lindau disease. Cancer Res 2002 Jul 1;62(13):3803-11.

(191) Ricketts C, Zeegers MP, Lubinski J, Maher ER. Analysis of germline variants in CDH1, IGFBP3, MMP1, MMP3, STK15 and VEGF in familial and sporadic renal cell carcinoma. PLoS One 2009 Jun 24;4(6):e6037.

(192) Webster AR, Richards FM, MacRonald FE, Moore AT, Maher ER. An analysis of phenotypic variation in the familial cancer syndrome von Hippel-Lindau disease: evidence for modifier effects. Am J Hum Genet 1998 Oct;63(4):1025-35.

(193) Rattenberry E, Vialard L, Yeung A, Bair H, McKay K, Jafri M, et al. A comprehensive next generation sequencing-based genetic testing strategy to improve diagnosis of inherited pheochromocytoma and paraganglioma. J Clin Endocrinol Metab 2013 Jul;98(7):E1248-E1256.

(194) Kleinjan DA, van H, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 2005 Jan;76(1):8-32.

(195) Nickerson ML, Jaeger E, Shi Y, Durocher JA, Mahurkar S, Zaridze D, et al. Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. Clin Cancer Res 2008 Aug 1;14(15):4726-34.

(196) Coppin L, Grutzmacher C, Crepin M, Destailleur E, Giraud S, Cardot-Bauters C, et al. VHL mosaicism can be detected by clinical next-generation sequencing and is not restricted to patients with a mild phenotype. Eur J Hum Genet 2014 Sep;22(9):1149-52.

(197) Wu P, Zhang N, Wang X, Li T, Ning X, Bu D, et al. Mosaicism in von Hippel-Lindau disease with severe renal manifestations. Clin Genet 2013 Dec;84(6):581-4.

(198) Smith MJ, Beetz C, Williams SG, Bhaskar SS, O'Sullivan J, Anderson B, et al. Germline mutations in SUFU cause Gorlin syndrome-associated childhood medulloblastoma and redefine the risk associated with PTCH1 mutations. J Clin Oncol 2014 Dec;32(36):4155-61.

(199) Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 2008 Nov;29(11):1282-91.

(200) Bell JA, Bodmer D, Sistermans E, Ramsden SC. Practice guidelines for the Interpretation and Reporting of Unclassified Variants (UVs) in Clinical Molecular Genetics. 22-10-2007.
Ref Type: Online Source

(201) Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. Hum Mutat 2008 Nov;29(11):1265-72.

(202) Tavtigian SV, Greenblatt MS, Goldgar DE, Boffetta P. Assessing pathogenicity: overview of results from the IARC Unclassified Genetic Variants Working Group. Hum Mutat 2008 Nov;29(11):1261-4.

(203) Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature 2010 Oct 28;467(7319):1061-73.

(204) Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012 Nov 1;491(7422):56-65.

(205) 1000 genomes consortium. 1000 genomes: A Deep Catalog of Human Genetic Variation. 2015.
Ref Type: Online Source

(206) Exome Variant Server. NHLBI GO Exome Sequencing Project (ESP). 2015. Seattle, WA.
Ref Type: Online Source

(207) Exome Aggregation Consortium (ExAC). Exome Aggregation Consortium (ExAC). 2015.
Ref Type: Online Source

(208) Veltman JA, Brunner HG. De novo mutations in human genetic disease. Nat Rev Genet 2012 Jul 18;13(8):565-75.

(209) Strachan T, Read A. Human Molecular Genetics. 4th ed. Garland Science; 2010.

(210) Rathmell WK, Chen S. VHL inactivation in renal cell carcinoma: implications for diagnosis, prognosis and treatment. Expert Rev Anticancer Ther 2008 Jan;8(1):63-73.

(211) Gill AJ, Benn DE, Chou A, Clarkson A, Muljono A, Meyer-Rochow GY, et al. Immunohistochemistry for SDHB triages genetic testing of SDHB, SDHC, and SDHD in paraganglioma-pheochromocytoma syndromes. Hum Pathol 2010 Jun;41(6):805-14.

(212) van Nederveen FH, Gaal J, Favier J, Korpershoek E, Oldenburg RA, de Bruyn EM, et al. An immunohistochemical procedure to detect patients with paraganglioma and phaeochromocytoma with germline SDHB, SDHC, or SDHD gene mutations: a retrospective and prospective analysis. Lancet Oncol 2009 Aug;10(8):764-71.

(213) Castelblanco E, Santacana M, Valls J, de CA, Cascon A, Robledo M, et al. Usefulness of negative and weak-diffuse pattern of SDHB immunostaining in assessment of SDH mutations in paragangliomas and pheochromocytomas. Endocr Pathol 2013 Dec;24(4):199-205.

(214) Frankel A. Formalin fixation in the '-omics' era: a primer for the surgeon-scientist. ANZ J Surg 2012 Jun;82(6):395-402.

(215) Tosi M, Stamm S, Baralle D. RNA splicing meets genetic testing: detection and interpretation of splicing defects in genetic diseases. Eur J Hum Genet 2010 Jun;18(6):737-8.

(216) Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 2002 Apr;3(4):285-98.

(217) Jian X, Boerwinkle E, Liu X. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. Genet Med 2014 Jul;16(7):497-503.

(218) Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 2008 Nov;29(11):1304-13.

(219) Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. Hum Mutat 2012 Aug;33(8):1228-38.

(220) Baralle D, Lucassen A, Buratti E. Missed threads. The impact of pre-mRNA splicing defects on clinical practice. EMBO Rep 2009 Aug;10(8):810-6.

(221) Messiaen LM, Callens T, Mortier G, Beysen D, Vandenbroucke I, Van RN, et al. Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. Hum Mutat 2000;15(6):541-55.

(222) National Genetics Reference Laboratory Manchester. Missense Prediction Tool Catalogue. 5-8-2014. 7-8-2015.
Ref Type: Online Source

(223) Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4(7):1073-81.

(224) Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 2006 Apr;43(4):295-305.

(225) Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 2011 Sep 1;39(17):e118.

(226) Thomas PD, Kejariwal A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. Proc Natl Acad Sci U S A 2004 Oct 26;101(43):15398-403.

(227) Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One 2012;7(10):e46688.

(228) Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010 Apr;7(4):248-9.

(229) Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 2006 Nov 15;22(22):2729-34.

(230) Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 2014 Apr;11(4):361-2.

(231) Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 2009 Nov 1;25(21):2744-50.

(232) Bao L, Zhou M, Cui Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. Nucleic Acids Res 2005 Jul 1;33(Web Server issue):W480-W482.

(233) Niroula A, Urolagin S, Vihinen M. PON-P2: prediction method for fast and reliable identification of harmful variants. PLoS One 2015 Feb 3;10(2):e0117380.

(234) Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res 2007;35(11):3823-35.

(235) Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 2009 Aug;30(8):1237-44.

(236) Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. BMC Genomics 2013;14 Suppl 3:S2. doi: 10.1186/1471-2164-14-S3-S2. Epub@2013 May 28.:S2-14.

(237) Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol 2014 Jan;10(1):e1003440.

(238) Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet 2011 Apr 8;88(4):440-9.

(239) Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat 2011 Apr;32(4):358-68.

(240) Couch FJ, Rasmussen LJ, Hofstra R, Monteiro AN, Greenblatt MS, de WN. Assessment of functional effects of unclassified genetic variants. Hum Mutat 2008 Nov;29(11):1314-26.

(241) Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 1987 Sep 11;15(17):7155-74.

(242) Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 2004;11(2-3):377-94.

(243) Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J Comput Biol 1997;4(3):311-23.

(244) Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res 2001 Mar 1;29(5):1185-90.

(245) Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009 May;37(9):e67.

(246) Ni Y, Zbuk KM, Sadler T, Patocs A, Lobo G, Edelman E, et al. Germline mutations and variants in the succinate dehydrogenase genes in Cowden and Cowden-like syndromes. Am J Hum Genet 2008 Aug;83(2):261-8.

(247) Ni Y, He X, Chen J, Moline J, Mester J, Orloff MS, et al. Germline SDHx variants modify breast and thyroid cancer risks in Cowden and Cowden-like syndrome via FAD/NAD-dependant destabilization of p53. Hum Mol Genet 2012 Jan 15;21(2):300-10.

(248) Nahorski MS, Reiman A, Lim DH, Nookala RK, Seabra L, Lu X, et al. Birt Hogg-Dube syndrome-associated FLCN mutations disrupt protein stability. Hum Mutat 2011 Aug;32(8):921-9.

(249) Ollila S, Dermadi BD, Jiricny J, Nystrom M. Mechanisms of pathogenicity in human MSH2 missense mutants. Hum Mutat 2008 Nov;29(11):1355-63.

(250) Bayley JP, Taschner PE. TCA Cycle Gene Mutation Database. 2015.
Ref Type: Online Source

(251) Bayley JP. Succinate dehydrogenase gene variants and their role in Cowden syndrome. Am J Hum Genet 2011 May 13;88(5):674-5.

(252) Lendvai N, Toth M, Valkusz Z, Beko G, Szucs N, Csajbok E, et al. Over-representation of the G12S polymorphism of the SDHD gene in patients with MEN2A syndrome. Clinics (Sao Paulo) 2012;67 Suppl 1:85-9.:85-9.

(253) Montani M, Schmitt AM, Schmid S, Locher T, Saremaslani P, Heitz PU, et al. No mutations but an increased frequency of SDHx polymorphisms in patients with sporadic and familial medullary thyroid carcinoma. Endocr Relat Cancer 2005 Dec;12(4):1011-6.

(254) Abbs S, Martin H, Brugger K, Rodger F, Littleboy R, Delon I, et al. Bridging the gap between sequencing gene panels and whole exomes for clinical diagnosis. 2015.

(255) Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg EJ, Mensenkamp AR, et al. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. Hum Mutat 2013 Dec;34(12):1721-6.

(256) Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2010 Jan;42(1):30-5.

(257) Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med 2013 Oct 17;369(16):1502-11.

(258) Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA 2014 Nov 12;312(18):1880-7.

(259) Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 2014 Nov 12;312(18):1870-9.

(260) Sawyer SL, Hartley T, Dyment DA, Beaulieu CL, Schwartzentruber J, Smith A, et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. Clin Genet 2015 Aug 18;10.

(261) Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet 2012 May;20(5):490-7.

(262) Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013 Jul;15(7):565-74.

(263) Wolf SM, Annas GJ, Elias S. Point-counterpoint. Patient autonomy and incidental findings in clinical genomics. Science 2013 May 31;340(6136):1049-50.

(264) PHG Foundation. Managing incidental and pertinent findings from WGS in the 100,000 Genomes Project. 1-5-2013. 8-9-2015.
Ref Type: Online Source

(265) Royal College of Pathologists of Australasia. Implementation of Massively Parallel Sequencing in Diagnostic Medical Genetic Testing. 23-1-2014. 8-9-2015.
Ref Type: Online Source

(266) Middleton A, Patch C, Wiggins J, Barnes K, Crawford G, Benjamin C, et al. Position statement on opportunistic genomic screening from the Association of Genetic Nurses and Counsellors (UK and Ireland). Eur J Hum Genet 2014 Aug;22(8):955-6.

(267) van El CG, Cornel MC, Borry P, Hastings RJ, Fellmann F, Hodgson SV, et al. Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. Eur J Hum Genet 2013 Jun;21(6):580-4.

(268) Hehir-Kwa JY, Claustres M, Hastings RJ, van Ravenswaaij-Arts C, Christenhusz G, Genuardi M, et al. Towards a European consensus for reporting incidental findings during clinical NGS testing. Eur J Hum Genet 2015 Jun 3;10.

(269) Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. Genet Med 2014 Feb;16(2):176-82.

(270) Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genet Med 2015 Jun;17(6):444-51.

(271) Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nat Rev Genet 2013 Jun;14(6):415-26.

(272) Genomics England. The 100,000 Genomes Project. 9-9-2015. 9-9-2015.
Ref Type: Online Source

Chapter 9 Peer-reviewed publications

9.1 Publication of work performed for part of Chapter 3

JCEM ONLINE

Advances in Genetics—Endocrine Research

A Comprehensive Next Generation Sequencing–Based Genetic Testing Strategy To Improve Diagnosis of Inherited Pheochromocytoma and Paraganglioma

Eleanor Rattenberry, Lindsey Vialard, Anna Yeung, Hayley Bair, Kirsten McKay, Mariam Jafri, Natalie Canham, Trevor R. Cole, Judit Denes, Shirley V. Hodgson, Richard Irving, Louise Izatt, Márta Korbonits, Ajith V. Kumar, Fiona Lalloo, Patrick J. Morrison, Emma R. Woodward, Fiona Macdonald, Yvonne Wallis, and Eamonn R. Maher*

Context: Pheochromocytomas and paragangliomas are notable for a high frequency of inherited cases, many of which present as apparently sporadic tumors.

Objective: The objective of this study was to establish a comprehensive next generation sequencing (NGS)–based strategy for the diagnosis of patients with pheochromocytoma and paraganglioma by testing simultaneously for mutations in *MAX*, *RET*, *SDHA*, *SDHB*, *SDHC*, *SDHD*, *SDHAF2*, *TMEM127*, and *VHL*.

Design: After the methodology for the assay was designed and established, it was validated on DNA samples with known genotype and then patients were studied prospectively.

Setting: The study was performed in a diagnostic genetics laboratory.

Patients: DNA samples from 205 individuals affected with adrenal or extraadrenal pheochromocytoma/head and neck paraganglioma (PPGL/HNPGL) were analyzed. A proof-of-principle study was performed using 85 samples known to contain a variant in 1 or more of the genes to be tested, followed by prospective analysis of an additional 120 samples.

Main Outcome Measures: We assessed the ability to use an NGS-based method to perform comprehensive analysis of genes implicated in inherited PPGL/HNPGL.

Results: The proof-of-principle study showed that the NGS assay and analysis gave a sensitivity of 98.7%. A pathogenic mutation was identified in 16.6% of the prospective analysis cohort of 120 patients.

Conclusions: A comprehensive NGS-based strategy for the analysis of genes associated with predisposition to PPGL and HNPGL was established, validated, and introduced into diagnostic service. The new assay provides simultaneous analysis of 9 genes and allows more rapid and cost-effective mutation detection than the previously used conventional Sanger sequencing–based methodology. (*J Clin Endocrinol Metab* 98: E1248–E1256, 2013)

9.2 Publication of work performed in addition to thesis

Research | F Haller et al. | SDHC methylation in Carney triad | 21:4 | 567–577

Aberrant DNA hypermethylation of *SDHC*: a novel mechanism of tumor development in Carney triad

Florian Haller, Evgeny A Moskalev, Fabio R Faucz[1], Sarah Barthelmeß, Stefan Wiemann[2], Matthias Bieg[3], Guillaume Assie[4,5], Jerome Bertherat[4,5], Inga-Marie Schaefer[6,†], Claudia Otto[7], Eleanor Rattenberry[8], Eamonn R Maher[8,9], Philipp Ströbel[6], Martin Werner[7], J Aidan Carney[10], Arndt Hartmann, Constantine A Stratakis[1] and Abbas Agaimy

Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Krankenhausstraße 8-10, D-91054 Erlangen, Germany
[1]Program on Developmental Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA
[2]Division Molecular Genome Analysis [3]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany
[4]Institut Cochin, INSERM U1016, CNRS UMR 8104, Université Paris Descartes, Sorbonne Paris Cité, Paris, France
[5]Department of Endocrinology, Referal Center for Rare Adrenal Diseases, Assistance Publique Hôpitaux de Paris, Hôpital Cochin, Paris, France
[6]Institute of Pathology, University Medical Center, Georg-August University, Göttingen, Germany
[7]Institute of Pathology, University Hospital, Albert-Ludwigs University Freiburg, Freiburg, Germany
[8]School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, Centre for Rare Diseases and Personalised Medicine, Birmingham Women's Hospital, University of Birmingham and West Midlands Regional Genetics Service, Birmingham, UK
[9]Department of Medical Genetics, University of Cambridge, Cambridge CB2 0QQ, UK
[10]Laboratory Medicine and Pathology, Emeritus Staff, Mayo Clinic, Rochester, Minnesota, USA
[†]I-M Schaefer is now at Department of Pathology, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts, USA

Correspondence should be addressed to F Haller or C A Stratakis
Emails
florian.haller@uk-erlangen.de or
stratakc@ccl.nichd.nih.gov

Endocrine-Related Cancer

Abstract

Carney triad (CT) is a rare condition with synchronous or metachronous occurrence of gastrointestinal stromal tumors (GISTs), paragangliomas (PGLs), and pulmonary chondromas in a patient. In contrast to Carney–Stratakis syndrome (CSS) and familial PGL syndromes, no germline or somatic mutations in the succinate dehydrogenase (*SDH*) complex subunits *A*, *B*, *C*, or *D* have been found in most tumors and/or patients with CT. Nonetheless, the tumors arising among patients with CT, CSS, or familial PGL share a similar morphology with loss of the SDHB subunit on the protein level. For the current study, we employed massive parallel bisulfite sequencing to evaluate DNA methylation patterns in CpG islands in proximity to the gene loci of all four SDH subunits. For the first time, we report on a recurrent aberrant dense DNA methylation at the gene locus of *SDHC* in tumors of patients with CT, which was not present in tumors of patients with CSS or PGL, or in sporadic GISTs with *KIT* mutations. This DNA methylation pattern was correlated to a reduced mRNA expression of *SDHC*, and concurrent loss of the SDHC subunit on the protein level. Collectively, these data suggest epigenetic inactivation of the *SDHC* gene locus with

Key Words
- Carney triad
- SDH complex
- SDHC
- GIST
- paraganglioma

http://erc.endocrinology-journals.org
DOI: 10.1530/ERC-14-0254

Published by Bioscientifica Ltd.

Heterogeneous Genetic Background of the Association of Pheochromocytoma/Paraganglioma and Pituitary Adenoma: Results From a Large Patient Cohort

Judit Dénes, Francesca Swords, Eleanor Rattenberry, Karen Stals, Martina Owens, Treena Cranston, Paraskevi Xekouki, Linda Moran, Ajith Kumar, Christopher Wassif, Naomi Fersht, Stephanie E. Baldeweg, Damian Morris, Stafford Lightman, Amar Agha, Aled Rees, Joan Grieve, Michael Powell, Cesar Luiz Boguszewski, Pinaki Dutta, Rajesh V. Thakker, Umasuthan Srirangalingam, Chris J. Thompson, Maralyn Druce, Claire Higham, Julian Davis, Rosalind Eeles, Mark Stevenson, Brendan O'Sullivan, Phillipe Taniere, Kassiani Skordilis, Plamena Gabrovska, Anne Barlier, Susan M. Webb, Anna Aulinas, William M. Drake, John S. Bevan, Cristina Preda, Nadezhda Dalantaeva, Antônio Ribeiro-Oliveira Jr, Isabel Tena Garcia, Galina Yordanova, Violeta Iotova, Jane Evanson, Ashley B. Grossman, Jacqueline Trouillas, Sian Ellard, Constantine A. Stratakis, Eamonn R. Maher, Federico Roncaroli, and Márta Korbonits*

Context: Pituitary adenomas and pheochromocytomas/paragangliomas (pheo/PGL) can occur in the same patient or in the same family. Coexistence of the two diseases could be due to either a common pathogenic mechanism or a coincidence.

Objective: The objective of the investigation was to study the possible coexistence of pituitary adenoma and pheo/PGL.

Design: Thirty-nine cases of sporadic or familial pheo/PGL and pituitary adenomas were investigated. Known pheo/PGL genes (*SDHA-D, SDHAF2, RET, VHL, TMEM127, MAX, FH*) and pituitary adenoma genes (*MEN1, AIP, CDKN1B*) were sequenced using next generation or Sanger sequencing. Loss of heterozygosity study and pathological studies were performed on the available tumor samples.

Setting: The study was conducted at university hospitals.

Patients: Thirty-nine patients with sporadic of familial pituitary adenoma and pheo/PGL participated in the study.

Outcome: Outcomes included genetic screening and clinical characteristics.

Results: Eleven germline mutations (five *SDHB*, one *SDHC*, one *SDHD*, two *VHL*, and two *MEN1*) and four variants of unknown significance (two *SDHA*, one *SDHB*, and one *SDHAF2*) were identified in the studied genes in our patient cohort. Tumor tissue analysis identified LOH at the *SDHB* locus in three pituitary adenomas and loss of heterozygosity at the *MEN1* locus in two pheochromocytomas. All the pituitary adenomas of patients affected by *SDHX* alterations have a unique histological feature not previously described in this context.

Conclusions: Mutations in the genes known to cause pheo/PGL can rarely be associated with pituitary adenomas, whereas mutation in a gene predisposing to pituitary adenomas (*MEN1*) can be associated with pheo/PGL. Our findings suggest that genetic testing should be considered in all patients or families with the constellation of pheo/PGL and a pituitary adenoma. (*J Clin Endocrinol Metab* 100: E531–E541, 2015)

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in U.S.A.

Received August 26, 2014. Accepted December 9, 2014.
First Published Online December 12, 2014

* Author Affiliations are shown at the bottom of the next page.

Abbreviations: AIP, aryl hydrocarbon receptor interacting protein; ER, endoplasmic reticulum; ERLEC1, ER lectin 1; FH, fumarate hydratase; FIPA, familial isolated PA; H&E, hematoxylin and eosin; LOH, loss of heterozygosity; MAX, MYC associated factor X; MEN, multiple endocrine neoplasia; MLPA, multiplex ligation-dependent probe amplification; NFPA, nonfunctioning pituitary adenoma; PAS, periodic acid-Schiff; PRL, prolactin; SDHA, succinate dehydrogenase subunit A; SDHAF2, succinate dehydrogenase complex assembly factor 2; SDHB, succinate dehydrogenase subunit B; SDHC, succinate dehydrogenase subunit C; SDHD, succinate dehydrogenase subunit D; PA, pituitary adenoma; pheo/PGL, pheochromocytoma/paraganglioma; RET, rearranged during transfection tyrosine kinase receptor; TMEM127, transmembrane protein 127; VHL, von Hippel-Lindau.