Exome Sequencing and Human Disease

The molecular characterisation of genetic disorders

by

Diana Maria Walsh

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

DOCTOR OF PHILOSOPHY

Institute of Biomedical Research

School of Clinical and Experimental Medicine

College of Medical and Dental Sciences

University of Birmingham

September 2015

Abstract

Since the completion of the human genome project in 2001, the field of genomics has advanced exponentially, largely in part to the introduction of next generation sequencing (NGS); a technique that has revolutionised the ways in which genetic disease is investigated. NGS enables the simultaneous sequencing of multiple reads in parallel, which provides researchers with the opportunity to interrogate vast numbers of candidate genes in order to establish the genetic eitiology and key components of disease. Exome sequencing in particular offers an efficient method to investigate disease, as the exomic regions make up 1% of the whole genome, but can contain up to 85% of functional variants responsible for disease. Next generation sequencing has been employed to investigate and identify the genetic cause of Acrocallosal syndrome (a rare autosomal recessive disorder). Exome sequencing was then also applied to investigate the genetic associations with both familial and sporadic pheochromocytomas and paragangliomas (neuroendocrine tumours). This study describes s the various applications, challenges and potential benefits that can be achieved by using exome sequencing as a tool to investigate rare autosomal recessive disorders in addition to more complex disorders including familial and sporadic cancer. This study aims to employ cutting edge technology to investigate human disease, in order to enhance current understandings of disease biology and pathogenesis. Through this, it is hoped that these findings may help to contribute to on-going efforts to develop novel therapeutic strategies and improve the clinical management of these disorders.

Acknowledgements

I would like to offer a sincere thank you to everyone from the molecular labs that have helped me over the years, especially Dean, Dewi and Malgosia for all of their scientific advice. I would also like to say thank you to my supervisor, Eamonn Maher for all of his guidance and for providing me with the amazing opportunity to carry out work in an exciting, cutting edge field. I would also like to say thank you to my second supervisor, Farida Latif, who offered me a lot of support throughout my time on this project. I would like to thank Jan, for putting up with the endless amount of pipette tips I used to generate, and for all of our lovely chats. I would also like to say a huge thank you my office, including Naomi Wake, for the incredible help she gave me throughout the project, Thoraia, Abdullah and Seley for our great conversations (also for the dates & Arabic coffee!). My partner in crime, and coffee bud, Amy- I enjoyed all of our chats in the office and will miss all of the fun and jokes we used to share (I still check my desk in the morning for pretend spiders!). I would like to dedicate this thesis and offer an incredibly special thank you to my mum and dad, who spent many Friday evenings in the Country Girl with me, listening to all of my genetics troubles over a glass of wine. To my dad for always having such a keen interest in everything that I do, I probably wouldn't have made it this far without your encouragement. Finally, to my new husband, Jamie, thank you so much for all of your support during the tough times, and for sticking with me through them! Coming home to you and Nova (our dog) used to make all the troubles seem so distant. I can't wait to finally spend some cosy evenings with you, without being surrounded by papers! You made it all worth it.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A	adenine	**mRNA**	messenger ribonucleic acid
bp	base pair(s)	**mTOR**	Mammalian Target of Rappamycin
BRCA	Breast Cancer Gene	**NF1**	Neurofibromatosis 1
C	cytosine	**ng**	nanogram
ccRCC	clear cell renal cell carcinoma	**NGS**	Next Generation Sequencing
cDNA	complementary DNA	**PCR**	polymerase chain reaction
CF	Cystic Fibrosis	**PHD**	Prolyl Hydroxylase Domain
CGH	Comparative Genomic Hybridization	**RB1**	Retinoblastoma Gene
COSMIC	Catalogue of Somatic Mutations in Cancer	**RCC**	Renal cell carcinoma
ddNTPs	dideoxynucleotides	**RNA**	ribonucleic acid
DIV	Deep Intronic Variant	**rpm**	revolutions per minute
DNA	deoxyribonucleic acid	**SDH**	Succinate Dehydrogenase
dNTP's	deoxyribonucleoside 5'-triphosphates	**SNP**	single nucleotide polymorphism
EDTA	ethylenediaminetetraacetic acid	**T**	thymine
ERF	ETS2 Repressor Factor	**TBE**	Tris-borate/EDTA
FH	Fumarate Hydratase	**TCGA**	The Cancer Genome Atlas
FISH	Fluorescent In Situ Hybridisaion	**Tris**	Tris(hydroxymethyl)aminomethane
G	guanine	**TSG**	Tumour Suppressor Gene
HIF	Hypoxia inducible factor	**U**	units
kb	kilobase pair	**UTR**	untranslated region
M	Molar	**UV**	ultraviolet

MAF	Minor Allele Frequency	**VHL**	Von-Hippel Lindau
Mb	megabase pair	**w/v**	weight/volume
ME	Malic Enzyme	**µl**	microlitre
MEN	Multiple Endocrine Neoplasia		
ml	millilitre		
mM	millimolar		

Chapter One:

INTRODUCTION

1.1 The Genetic Epidemiology of Inherited Disease

It has been 150 years since Gregor Mendel performed his unknowingly ground-breaking investigations into the hybridisation of pea plants. Through his observations and investigations into the patterns of heritability from one generation to the next, Mendel inadvertently formed the foundations and basis of modern day genetics. He managed to establish that alleles are inherited in pairs (one from each parent), and also that certain traits are inherited in a dominant fashion while others are recessive and remain 'hidden' until subsequent generations. Mendel also determined that the inheritance of one characteristic is not influenced by the inheritance of another. Through these observations, he managed to establish three main theories of inheritance; these are now known as the law of segregation, the law of independent assortment and the law of dominance (Mendel & Bateson 1865). These principles, although now often considered to be a vast oversimplification, remain the fundamental principles around which all genetics studies revolve around today.

1.1a Mendelian Diseases: Clinical Aspects

According to Mendel's principles, diseases can be classified into groups based on their mode of inheritance, including autosomal recessive disorders, autosomal dominant disorders, X-linked and Y-linked.

1.1a.i Autosomal Recessive Disorders

Autosomal recessive disorders refer to those that are caused by the inheritance of two mutant alleles for a particular disease gene. For example, if two parents are carriers of a pathogenic variant in a disease gene, their offspring will have a 50% chance of being born as an unaffected heterozygous carrier, 25% chance of being born an unaffected non-carrier, and a 25% chance of being born homozygous for both mutant alleles resulting in disease manifestation. Cystic fibrosis (CF), a disorder characterised by the secretion of thick mucus in the lungs and airways of affected individuals, is one of the most well recognized autosomal recessive disorders, and is known to affect approximately 70,000 individuals globally (Cutting 2014). The inheritance of two mutated copies (alleles) of the *CFTR* gene is required for the development of CF; although the degree of severity of the disorder is known to be variable.

1.1a.ii Autosomal Dominant Disorders

Autosomal dominant disorders manifest when only one mutant allele from a disease gene is inherited. For example, Huntington's disorder, a neurodegenerative disease, can manifest in individuals who have inherited a single pathogenic mutation in the *HTT* gene (Burgunder 2014). Affected individuals with a pathogenic mutation will have a 50% chance of passing their mutation on to any offspring.

Vast and rapid advances in the abilities of sequencing technologies have enabled researchers to apply Mendel's principles in order to elucidate the genetic landscape of many inherited disorders, and progress our understanding of the biological mechanisms involved in their pathogenesis. However, with such advancement in our abilities to sequence DNA and ascertain information regarding the human genome, has also come the realisation that the inheritance and development of genetic disease can be much more complex than Mendel originally believed. Concepts such as incomplete penetrance, variable expressivity, multi-gene traits, modifier genes and oligogenic inheritance are but a few genetic phenomena that can play roles in genetic disease. Advancements in our understanding of these concepts have shifted our perceptions of disease transmission in recent years, and it is now beginning to become apparent that an expanding number of diseases cannot be completely explained by simplistic Mendelian inheritance alone. Rather, it is more common that genetic diseases are the products of a convoluted and often highly individualised genetic web of interacting factors, that collectively contribute to the final expression of the clinical phenotype.

1.1a.iii Variable Penetrance and Variable Expressivity

In some disorders, mutations in the same gene can generate different clinical effects in different individuals. For example, certain carriers of a mutation may express the disease phenotype while others might not. In other cases, the phenotype may be expressed in all individuals, but there may be a high degree of variability between their clinical features.

These phenomena are referred to as incomplete penetrance and variable expressivity; both of which are likely to occur as a result of a unique combination of both genetic contributory factors and environmental exposures. As these factors are likely to be highly personalised, it is notoriously difficult to predict the likely clinical and phenotypic outcome of each carrier of a specific genotype.

1.1a.iv Penetrance

Penetrance can be defined as the proportion of carriers of a given genotype that express the associated characteristic phenotype (Zlotogora 2003). If a disease is described to have complete penetrance, this indicates that every carrier of a pathogenic mutation in the disease gene will always express the associated phenotype. For example, Neurofibromatosis type 1 is a highly penetrant disorder and almost all carriers of a pathogenic mutation in the *NF1* gene will express clinical features to a certain degree (K. Boyd, B. Korf, A. Theos 2009).

Conversely, a disease or gene is said to have incomplete or reduced penetrance when a proportion of carriers of a pathogenic mutation fail to express the associated characteristics (Shawky 2014). An example of this can be found in carriers of mutations in the *BRCA1* and *BRCA2* genes. All carriers have an increased lifetime risk of developing cancer, and although the majority do develop cancer at some stage in their lives, some carriers do not (Antoniou *et al* 2004, Cooper et al. 2013). This incomplete

penetrance is likely to be due to a complex interplay of both genetic and environmental factors; however the complete mechanisms that give rise to these situations remains unknown. For this reason, it is impossible to predict which *BRCA1/BRCA2* carriers will develop cancers and which will not; although, this is an area of research where further clarification could provide an extensive degree of clinical benefits.

1.1a.v Pseudoincomplete Penetrance

In some cases, non-penetrance can be incorrectly assumed in an individual due to an incomplete clinical examination or delayed onset of the phenotypic expression (e.g. age-dependent onset of cancers in *BRCA1/BRCA2* mutation carriers) (Shawky 2014). In such cases, this is referred to as pseudoincomplete penetrance. This can also arise when incomplete penetrance is wrongly assumed for a patient that is in-fact a mosaic carrier for a mutation. Thus in individuals with germline mosaicism, some of their gametes carry a mutation in a disease gene, and although not clinically affected themselves, they may have multiple affected children (Biesecker & Spinner 2013). In this way, it appears that the disease may be non-penetrant, but it is really due to the fact that the healthy parent does not carry the mutation in their somatic cells.

1.1a.vi Variable Expressivity

In some cases, although a disorder may be highly penetrant and manifest symptoms in most carriers, there can be a high degree of variability between the clinical features,

degree of severity and age of onset between patients (Cooper et al. 2013). This concept is described as variable expressivity, and in some disorders, such as CF, even high degrees of intrafamilial variation can be observed. CF patients exhibit wide and variable degrees of severities, can manifest an array of different physiological complications and can unpredictably different lengths of survival (Cutting 2014). This phenotypic variability can also occur in patients who harbour identical disease genotypes, which indicates that the phenotypic differences in these individuals must be due to environmental or genetic influences that are independent from the original disease mutation. Examples such as these reinforce the notion that even in monogenic disorders with apparent Mendelian inheritance, there can still be a wide variety of genetic contributory factors involved in determining multiple aspects to the disease.

1.1a.vii Mechanisms that Give Rise to Variable Penetrance and Expressivity

Although the complete mechanisms giving rise to variable penetrance and expressivity have not been completely elucidated, some of the contributory factors have been identified which can explain a proportion of some disorders. In some instances, it can be simple to comprehend the mechanisms giving rise to variable penetrance and expressivity. For example, it is understandable why male carriers of mutations in *BRCA2* may have a lower lifetime risk of acquiring breast cancer (6%), than female carriers (86%) (Feldman *et al* 2014). In other less obvious cases, the intragenic location of a mutation can affect disease penetrance (e.g. the common pathogenic mutation,

p.Phe508del in *CFTR* is highly penetrant, while the alteration p.Arg117His is associated with reduced penetrance) (Cutting, 2015). The type of mutation in itself can also have an effect; in general more deleterious types of mutations (nonsense and frameshifts) are associated with higher penetrance and more severe phenotypes than missense mutations as they are more likely to disrupt protein function. These influences are fairly simple to discern, particularly as clear, singular "cause and effect" relationships can be established. However, in many instances there are more complex factors that can influence expressivity and penetrance.

Digenic and oligogenic inheritance refers to the situation in which more than one gene contributes to a disease phenotype. These types of inheritance pattern can present considerable challenges when trying to establish genotype-phenotype relationships, and they can also be largely responsible for variations in expressivity and penetrance in some disorders. In Nephronophthisis (NPHP) for example, 6 disease genes have been identified. In addition, a case of digenic inheritance has been documented, in which one affected individual carried two separate mutations in two different disease genes that both contributed individually to disease manifestation (Hoefele et al. 2007).

In disorders that exhibit oligogenic and digenic patterns of inheritance, the resulting clinical phenotype that develops is dependent on which particular mutation or combination of mutations is inherited. The presence of modifier genes (although not

required for disease manifestation) can alter the effects of an existing mutation, either exacerbating or reducing their physiological consequences. For example, a recent study in Neurofibromatosis (NF1) patients managed to establish that several common polymorphisms were enriched in patients with higher counts of "café-au-lait" macules (hyperpigmented patches on the skin); through this, they were able to establish that these particular polymorphisms were contributing directly to a specific aspect of the variable expressivity seen in this disorder (Pemov et al. 2014). Distinguishing modifying alterations (which can also be present in healthy individuals with no observable effects) from a vast number of additional background variants can be a mammoth task; however studies like the one carried out by Pemov et al. 2014 demonstrate that the genetic influences that modify disease phenotypes can be elucidated with careful and thorough investigative approaches.

Both somatic and germline mosaicism can have an effect on the relative penetrance and expressivity of some disorders. Somatic mosaicism occurs when an individual harbours a different genotype in a small proportion of their cells, while germline mosaicism describes the situation in which some of an individual's gametes harbour a different genotype to their own (Biesecker & Spinner 2013). For example, in somatic mosaicism, the disease phenotype may be less severe due to the fact that the pathogenic mutation is only present in a reduced proportion of cells. In germline mosaicism, as the carrier parent may not exhibit clinical features of the disorder as the mutation is only present in

some of their gametes, and so the disease can appear to have reduced penetrance; however, any offspring that inherit this mutation will harbour it in all of their cells, resulting in disease manifestation.

Of all the factors influencing expressivity and penetrance in disease, environmental modifiers are arguably the most challenging to determine. This is largely because environmental exposures are highly individualised and their influential consequences can also have variable effects in different individuals. Furthermore, environmental modifiers cannot easily be replicated in research, and it is therefore difficult to reliably establish direct associations. Nevertheless, some clearly established associations have been made between environmental stimuli and disease manifestations. For example, it is commonly accepted that repeated exposures to certain carcinogens, such as UV raditation and cigarette smoke can introduce genetic mutations that will convey an increased susceptibility to cancer.

Another clear example of how environmental factors can be modified to alter a disease phenotype can be found in individuals affected with phenylketonuria; a rare genetic condition in which sufferers cannot metabolise phenylalanine. In this disease, new-born babies with high levels of phenylalanine can adopt a special phenylalanine-free diet which enables them to prevent the on-set of the severe effects of this disorder, including brain damage (Cooper et al. 2013). In this way, a clear modification to the environment

of a carrier can directly influence and reduce the penetrance of their disorder. The examples outlined here present clear-cut, simplistic associations; however in many disorders, combinations of many different environmental factors can interact with many, (often unknown), genetic modifiers, collectively resulting in an extremely complex and unpredictable genetic landscape that can present vast clinical challenges with regard to disease management.

1.2 Identification of the Genetic Basis of Disease

There are many types of genetic alteration that can occur within the genome that can be associated with disease, including base level mutations, copy number alterations, amplifications, deletions and translocations. For this reason, genetic disorders can loosely be categorised into 3 separate groups depending on their mode of pathogenesis; single gene or Mendelian disorders, chromosomal disorders and multifactorial disorders.

The use of cytogenetic and molecular laboratory techniques has enabled the discovery of many of these types of alterations at both the base pair and chromosomal level. The discoveries of such alterations and their roles in disease have not only enabled clear clarifications of many genes and pathways, but have provided novel therapeutic and diagnostic opportunities that have had major patient benefits.

The field of cytogenetics aims to identify the genetic causes of disease on a larger scale by enabling the detection of both structural and numerical chromosomal alterations, while molecular methods act on a smaller scale to detect single base substitutions and small indels at the DNA level.

1.2a Cytogenetics

Traditional cytogenetic approaches such as fluorescent in situ hybridization (FISH) and comparative genomic hybridization (CGH) have played major roles in the discovery of aberrations related to cancer. FISH is a technique that uses fluorescently labeled probes to bind to regions of interest on a chromosome, so that they physical chromosomal location of that region may be visualised. CGH is a technique used to identify copy number variation and unbalanced chromosomal anomalies by comparing and observing differences between two DNA samples; tumour and corresponding normal DNA for example. CGH is often used to identify specific chromosomal regions that are frequently lost or gained during tumour progression; information that can then be used to guide clinical decisions.

Chromosomal rearrangements are a common occurrence often observed in leukaemias and lymphomas, and have been shown to lead to the formation of gene fusions, or hybrid genes with oncogenic characteristics (Michels et al. 2007). The use of cytogenetic techniques led to the identification of the chromosomal locations of the genes associated

with the occurrence of this oncogenic event, which exposed novel therapeutic targets, and facilitated the development of the tyrosine-kinase inhibitor, Imatinib (Mitelman et al. 2007; Michels et al. 2007). This had a hugely positive impact on many patient lives, as the drug enabled a less harmful and more efficient means of treatment for chronic myeloid leukaemia. Discoveries such as these highlight the potential benefits that can be achieved through the identification and clarification of cancer genes and pathways implicated in the pathogenesis of cancer.

1.2b Molecular Methods for the Investigation of Genetic Disease

Mutations at the DNA level can arise in many different forms, including single nucleotide alterations as well as insertions and deletions. These alterations can have different functional impacts based on the changes they exert within the gene, and their subsequent effects on the translation of the protein encoded. To succeed in identifying these alterations and their roles in disease, there have been a number of molecular techniques that have been developed over the years.

1.2b.i Positional Cloning

In the days prior to the completion of the human genome project, the traditional method of gene detection was performed through positional cloning. This technique is used to identify the position of a disease gene through comparative studies in large family

pedigrees containing both affected and unaffected individuals. By using known polymorphic genetic markers, the region of the genome thought to contain the variant of interest can be identified based on which genetic markers are consistently co-inherited with disease phenotypes (Little 2005).

1.2b.ii Candidate Gene Approach

The candidate gene approach refers to the selection of a potential disease gene based on current understandings of the disease. This technique requires a *priori* knowledge in order to make an educated and informed prediction of which genes may be associated with disease. For example, a list of candidate genes can be selected based on their functional relevance (i.e. involvement in relevant biological pathways). Genes can also be selected using a position-dependent approach, in which candidate genes are selected based on prior linkage information (although this may not be effective for genetically heterogeneous disorders). Candidate genes can also be selected using information obtained from comparative genomics, basing decisions on the notion that mutations in genes that are highly conserved across many species are more likely to be detrimental.

Although many of these approaches seem logical, the candidate gene approach has been criticised by many for being a highly subjective technique that can be prone to bias during gene selection. Nevertheless, it remains a technique that has had some successes, and continues to be developed (Pasche & Yi 2010).

More recently, computational tools have been developed to combine an array of web-based information from public databases in order to prioritise candidate genes. This strategy is known as the digital candidate gene approach (DigiCGA); some of the programs developed including ENDEAVOR and SUSPECTS among others, utilise biological ontology data to compile their list of candidates (Zhu & Zhao 2007). They can also filter lists of candidates based on their expression in certain tissues, pathway involvements, known interacting partners and known associations with similar diseases. The programs can then use all of this information to perform complex statistical calculations in order to form prioritised lists of candidate genes. These programs are still in their infancy, and their track history of success has been highly variable; particularly with regard to more complex and multifactorial diseases (Zhu & Zhao 2007).

As with many scientific procedures, the strengths of the candidate gene approach can be found when using a combination of all of these techniques to identify candidates. In this way, the advantages of each strategy can be exploited in the hope that a less biased selection of candidate genes may be achieved. These approaches are particularly useful when analysing vast numbers of candidate genes such as that created by next generation sequencing.

1.3 Cancer as a Genetic Disease

Cancer develops following an accumulation of specific genetic alterations that enable a cell to become unresponsive to normal regulatory mechanisms. These alterations initiate uncontrolled growth and proliferation in the affected cells. Cells which have acquired these abilities subsequently undergo positive selection, resulting in the creation of small populations of cellular clones. As these clones grow and multiply they can ultimately lead to the formation of a tumour (Tabassum Kornelia Polyak, 2015).

The first coherent descriptions of cancer as a genetic disease came from studies performed by Theodor Boveri, who was the first to notice that different chromosomes can transmit different inheritable aspects that can have a unique influence on cell fate. Boveri later suggested that abnormal chromosomes contribute to tumour growth when he noticed that some cells with a "particular incorrect combination of chromosomes" resulted in the generation of a cell which displayed the ability to grow without restriction or limits, and that this abnormality was being passed to the cell's progeny (Boveri 1914). Even though the concept of a gene had not yet been described by 1914, Boveri accurately suggested that 'chromosome' imbalances or abnormalities were leading to tumour growth; this was the very first description of cancer occurring and progressing as a direct result of a genetic mutation.

Genes associated with cancer can be loosely classified into two groups; oncogenes (genes involved in promoting cell growth) and tumour suppressor genes (protective genes that can inhibit cell growth, initiate apoptosis and repair DNA). Mutations leading to cancer often provide a specific growth advantage either by causing activation or increased expression of oncogenic protein products, or by causing loss of function and reduced expression of proteins encoded by tumour suppressor genes.

1.3a Oncogenes

The first retroviral oncogene was discovered in the Rous Sarcoma Virus, which had been isolated from sarcomas in chickens by Peyton Rous in 1911 (Rous 1911). Later studies performed by Michael Bishop revealed that oncogenes carried by viruses were not actually viral genes, but mutated forms of genes that exist inherently in many organisms, including humans (Bishop 1989). The realisation was then made that cancer was different from other diseases in that it could arise in the absence of a foreign agent through specific alterations to normal cellular genes.

Non-mutated forms of oncogenes are known as proto-oncogenes; these genes are often involved in promoting cellular growth and proliferation. Therefore, in order to generate a pro-tumourigenic effect, they must acquire a gain-of-function alteration that causes constitutive activation or amplified expression of the encoded protein. These alterations

can occur through genetic mutation, gene amplifications and chromosomal rearrangements.

Activating mutations in human tumours often occur in the form of single base substitutions. Only very specific mutations can result in a gain-of-function, and therefore, cancer-related mutations in oncogenes are often found to cluster at, or within, specific regions of the gene encoding critical functional domains. These mutations can cause conformational changes within the essential regulatory domains, which then prevents their recognition and targeted degradation by E3 ubiquitin ligases; this is the mechanism by which activating *HIF2A* muations exert their tumourigenic effects in PCC/PGL tumours (Crona et al. 2013). Other mechanisms of gain-of-function can occur through activating mutations that enable the protein to activate itself, as is the case for activating *RET* mutations in multiple endocrine neoplasia (MEN); in these cases, *RET* mutations enable RET protein homodimerization, enabling auto-activation of the protein and subsequent initiation of downstream pathways (Donis-Keller et al 1993).

1.3b Tumour Suppressor Genes and Knudson's 2-hit Hypothesis

Tumour suppressor genes encode proteins that are protective to the cell, in that they can either repress cell growth, initiate apoptotic pathways or initiate DNA damage repair. Inactivation of a tumour suppressor gene through mutation, deletion or translocation can therefore lead to uncontrolled growth and proliferation, evasion of apoptosis or the

inability to repair damaged DNA; all of which are pro-tumourigenic characteristics (Hanahan & Weinberg 2011).

Some of the first descriptions of tumour suppressor genes in cancer occurred when Alfred Knudson released his famous "2-hit hypothesis" in 1971. While investigating retinoblastoma cases, Knudson made the observation that two rate-limiting events appeared to be necessary for disease manifestation; he subsequently postulated that in the inherited cases, one of the mutations had been passed through the germline and a second mutation was acquired during life in the somatic cells; he then reasoned that in non-hereditary cases, both mutational events had been acquired in somatic cells, possibly through the environment (Knudson 1971). Consistently with this, after the identification of the *RB1* TSG, molecular studies by other investigators revealed that the majority of those with familial or bilateral disease carried a germline mutation in the *Rb1* gene (the first hit), and their tumours showed loss of heterozygosity, confirming the presence of a second hit. The same types of investigations into non-inherited cases revealed that both of the "hits" had occurred somatically, as the patients did not carry any germline *Rb1* alterations. Knudson's 2-hit hypothesis therefore provides an explanation as to why the inheritance of a pathogenic mutation increases susceptibility to cancer, and also explains why those with inherited cancer often develop tumours at a young age.

1.3c Intratumoural Heterogeneity

It is well known that the cancer genome contains an extensive degree of genetic diversity; both between tumour types (intertumoural heterogenetity) and also within individual tumours (intratumoural heterogeneity). This is often a consequence of high levels of genomic instability, which increase the rates with which new mutations develop (Burrell et al. 2013). As tumourigenesis advances, vast numbers of new mutations can be acquired; however, not every new mutation will be present in every tumour cell. An increasing amount of evidence now suggests that contrary to prior belief, tumours are not simply clonal expansions from a single cell containing pro-tumourigenic mutations; but rather, tumours can contain many different clones, each with a unique set of tumour-related mutations, as displayed in figure 1 (Yates & Campbell 2014, McGranahan & Swanson 2015, Burrell et al. 2013). A study by Gerlinger et al. 2012, showed that from 9 different regions of a single kidney tumour, only 34% of the 118 different mutations identified were present in all cells from the 9 regions. This demonstrated that a high degree of subclonal diversity could exist within a single tumour, and that each clonal population could harbour very different and individualised genetic mutations (Gerlinger et al. 2012).

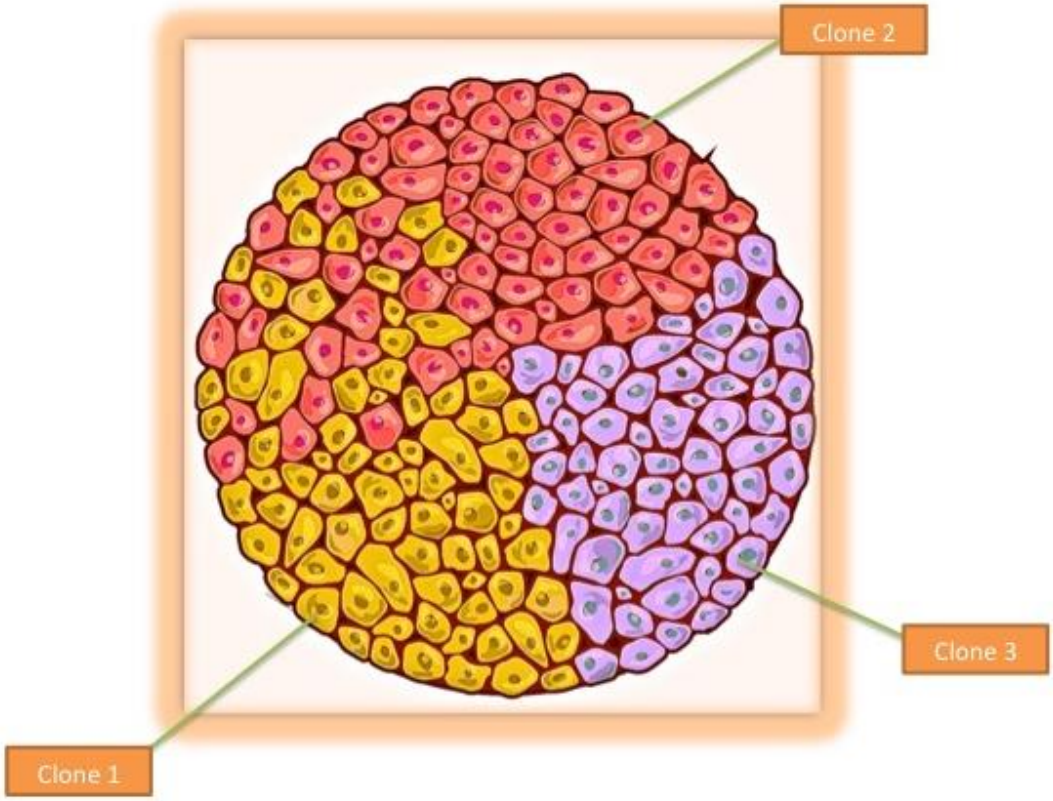


Figure 1. Multiclonal Populations in a Tumour Cell (adapted from Burrell et al. 2013)

Some of these clonal populations will harbour the key mutations that played an initiating role in tumour development, while other clones, which may have been acquired at a later stage, contain mutations that have played roles in driving tumour progression. These mutations that play an active role in tumourigenesis are known as drivers. These advantageous driver mutations will co-exist with numerous 'passenger' mutations (biologically neutral mutations that offer no selective growth advantage) and other

alterations that are also present in genomically normal non-cancer cells (Greenman et al. 2007; Stratton et al. 2009).

A major challenge in genomic studies of cancer is, primarily, to have the ability to identify the presence of potentially low-level mutations whose frequency may appear diluted due to the mosaic nature of tumour genetics; and secondly to distinguish between driver and passenger mutations in an environment of extensive intratumoural genetic heterogeneity.

1.4 Familial Cancer

Approximately 5-10% of cancers are associated with some form of inheritance pattern (Nagy et al. 2004). There are certain criteria that are strongly suggestive of a familial cancer syndrome; namely the presence of two or more relatives with the same cancer, several affected generations and the individuals have been diagnosed at a young age and/or have developed multiple primary tumours. The fact that there are families which display clear clusterings of similar cancers that fulfill most of the above criteria provides strong evidence that cancer can be a heritable disease. In the majority of cases, hereditary cancer syndromes are passed with an autosomal dominant inheritance pattern with incomplete penetrance.

In familial cancer syndromes, the inheritance of a pathogenic mutation provides an increased predisposition to a cancer; however, as cancer is a multistep process, the accumulation of further genetic events are required in order for tumourigenesis to occur. These additional genetic events are usually acquired through environmental exposures to carcinogens, or through lifestyle habits. Most familial cancer syndromes are associated with mutations in tumour suppressor genes (most commonly *P53*) (Vogelstein et al. 2013); however there are some familial cancer syndromes, such as MEN2 that are known to be associated with mutations in oncogenes (e.g. *RET*). Continued efforts to identify and characterise genes associated with familial cancer syndromes, will provide a wealth of preventative, diagnostic and therapeutic benefits that can be applied clinically to improve both the quality and length of life of cancer patients and their families.

1.5 Sporadic Cancer

Most cancers occur sporadically in the absence of any familial history. In these cases, the pathogenic mutations that are responsible for initiating tumourigenesis will have been acquired throughout the patient's life as opposed to inherited. Mutations can be acquired through many different stimuli, including exposure to carcinogens, such as harmful radiation, x-rays, UV-rays or through certain lifestyle habits including lack of exercise, high fat diets etc.

As previously described, cancer manifests following an accumulation of multiple genetic alterations; and therefore, unsurprisingly, the risks of cancer increase with age. Unlike familial cancer, where the successful identification of a susceptibility gene can offer relatives the opportunity to predict and prevent the onset of cancer, investigations into sporadic cancers usually aim to identify tumour-specific mutations that are absent from the germline DNA. The successful characterisation of a tumour driver's role in cancer progression may then help to identify potential molecular targets that can be exploited therapeutically. Furthermore, successful identification of mosaic mutations representative of smaller subclonal populations may provide information that can be used to combat the issues of chemotherapy resistance, by creating personalised, combinational therapies to target separate clones.

1.6 Next Generation Sequencing

Historically, the traditional method of investigating the genome was through the use of Sanger sequencing; a technique developed by Fredrick Sanger in 1977. Sanger sequencing involves base-specific termination of the DNA chain through the use of dideoxy nucleotides (ddNTPs). The ddNTPs lack the hydroxyl group necessary to elongate the DNA, and therefore terminate the chain each time one is incorporated, resulting in DNA fragments of varying lengths. These fragments are then separated

according to size by gel electrophoresis (Bayés et al. 2012). The Sanger method of sequencing was initially used with radioactively labelled ddNTPs that were detected using slab-gel electrophoresis. This method was later replaced by the dye-terminator method in which ddNTPs were instead fluorescently labelled and detected using capillary electrophoretic systems (Tucker et al. 2009). This revolutionised the field of genomics, and instigated the sequencing of the entire human genome, which initially took 13 years to complete, and cost approximately $2.7 billion (Zhang et al. 2011).

As sequencing technology continues to advance and the main companies compete for market share, costs are continually decreasing while throughput is increasing, resulting in a trend originally described to be similar to Moore's law; however since 2010, next generation sequencing technological trends have now even outpaced Moore's law, emphasizing the speed with which sequencing costs are falling while capabilities advance (figure 2). Within the last 50 years, the genomics field has advanced at such an astounding rate, that it has progressed from Watson and Crick's discovery of DNA to Illumina's most recent announcement that, with their new technology, they can sequence a full genome for $1000 (approximately £640) within a couple of days. This remarkable progression has now led on to the initiation of the next big genetic venture; The 100,000 Genomes Project. This is an enormous project that aims to sequence 100,000 genomes from 75,000 patients and 25,000 tumours (Siva 2015). The 100,000 genomes project is an undertaking that will provide an invaluable wealth of information with regard to

many genetic diseases, and propel the advances and achievements in the genetics field even further.

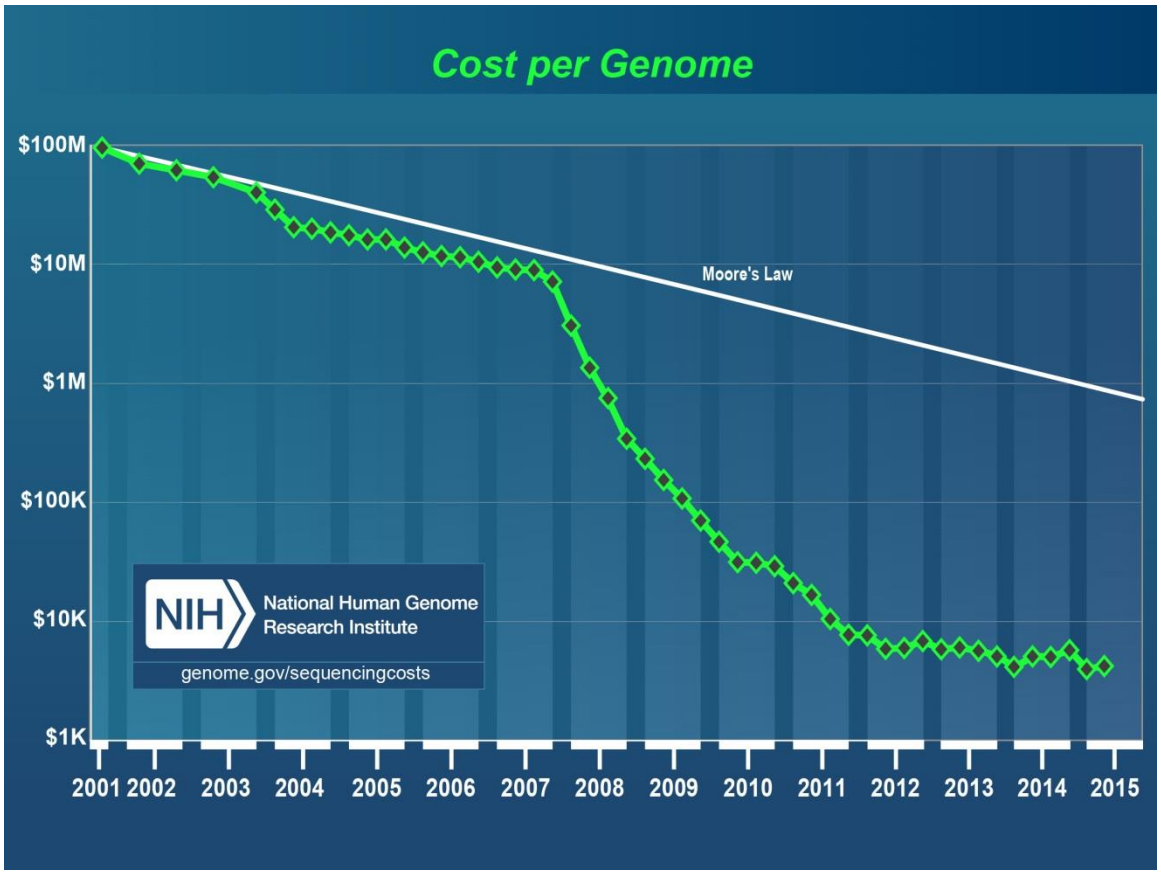


Figure 2. A graphical representation of the falling cost of genome sequencing in comparison to Moore's law. (as displayed from National Human Genome Research Institute, 2015)

Next generation sequencing has revolutionised the genomics field; it has generated an abundance of genetic information with regard to disease and has even enabled the identification of naturally occurring variants in plants that can be exploited to provide

growth advantages to certain types of crop. However, even though costs of sequencing continue to decrease rapidly, carrying out whole genome sequencing to a depth sufficient enough to identify disease-associated variants in large numbers of individuals still remains prohibitively expensive for most research projects. Therefore, with regard to disease gene discovery, the most commonly adopted approach involves the use of exome sequencing.

Exome sequencing only targets the coding regions of the genome; a region which makes up 1% of the genome, but contains approximately 85% of the functional variants responsible for disease (Majewski et al. 2011). Exome sequencing enables deeper sequencing, which therefore permits base calling with higher confidence levels. Furthermore, exome sequencing is more cost-effective, as a much smaller region is sequenced, which also means that the overall redundancy of information is greatly reduced, as WGS would generate larger quantities of variants of unknown significance.

1.6a Main Principles of Exome Sequencing

The basic concept of next generation sequencing is very similar to Sanger sequencing in that DNA polymerase is used to catalyse the incorporation of fluorescently labeled deoxynucleotides (dNTPs) during the extension of a strand of DNA from a template strand (Illumina 2015). Each of the fluorescently labeled dNTPs will emit a specific fluorophore during the sequencing process so that each base can be distinguished from

one another. Unlike Sanger sequencing, this process can be carried out in a massively parallel manner, enabling the analysis of multiple reads simultaneously. This 'sequencing by synthesis' method has become the most widely used technique across the NGS industry (Illumina 2015).

1.6b Exome Sequencing Process

The full process of exome sequencing can be broadly separated into 3 major stages: sample preparation and sequencing, raw data processing, and final data interpretation (figure 3).

In brief, the library preparation and sequencing stage includes the creation of a DNA library, capture of the exomic regions and targeted enrichment for the captured regions, followed by sequencing. The processing of the raw data then includes the alignment of the sequence to a reference genome followed by the use of *in silico* tools and programs to call variants. The final stage is to analyse the data and filter the variants in order to identify the variants of interest and interpret the findings.

There are numerous variations with regard to the different protocols, processes and analytical tools that can be applied when carrying out exome sequencing; therefore, this thesis will predominantly focus on the methods that are directly applicable to this study.

The specific methods used for the exome sequencing performed in this study can be found in section 2.3.

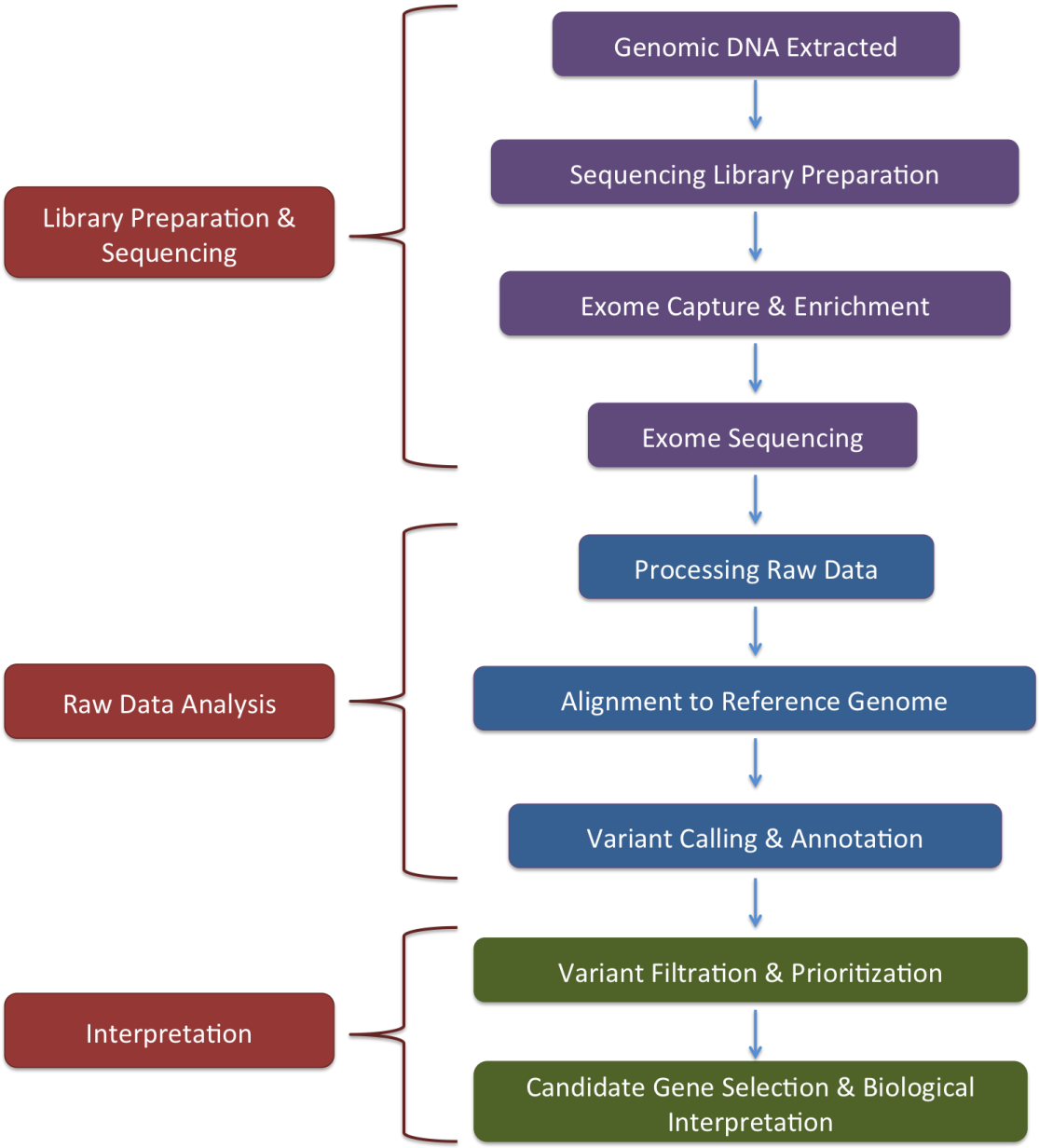


Figure 3. ***A Summary of the Main Stages Involved in the Process of Whole Exome Sequencing***

1.4c Library Preparation and Sequencing

Currently, the most prevalent method used to prepare a DNA library for exome sequencing is by a hybridisation-based enrichment process (Rykalina et al 2014). The main commercial providers of exome enrichment kits include Agilent, NimbleGen and Illumina; and although there are certain variable details between each kit, the fundamental principles of each protocol remain the same.

To begin with, genomic DNA is sheared into fragments of varying length, followed by the ligation of platform-specific adapters to the ends of the fragments. These steps are sometimes combined into a single step known as 'tagmentation', which enable DNA fragmentation and adapter ligation to occur in the same reaction; this greatly increases the efficiency of the library preparation process (Rykalina et al. 2014). The tagmentation method also has other advantages in that it minimizes the need for large concentrations of DNA, and can work with DNA concentrations as low as 50ng/µl. This is because the tagmentation process means that less DNA is lost at the fragmentation stage (Rykalina et al. 2014). The DNA library is then formed from the collection of adapter-ligated fragments; this library is then amplified by PCR and hybridised to biotinylated RNA baits, which have been specifically designed to target exomic regions. The hybridized exomic regions can then be pulled down efficiently through the use of streptavidin coated magnetic beads, which then allows unhybridized and intronic regions to be washed away. This library, now containing only exomic regions, can then be 'enriched'

through another PCR amplification step in order to generate a sufficient concentration for sequencing.

During sequencing, each fluorescently labeled dNTP added will emit a specific fluorophore; the intensity of which is proportional to the number of nucleotides added. This fluorescence data is then processed by the computer and imaged. The raw data produced by the sequencing machines is often received in a complex format, requiring a large amount of processing before it is ready to be analysed and interpreted.

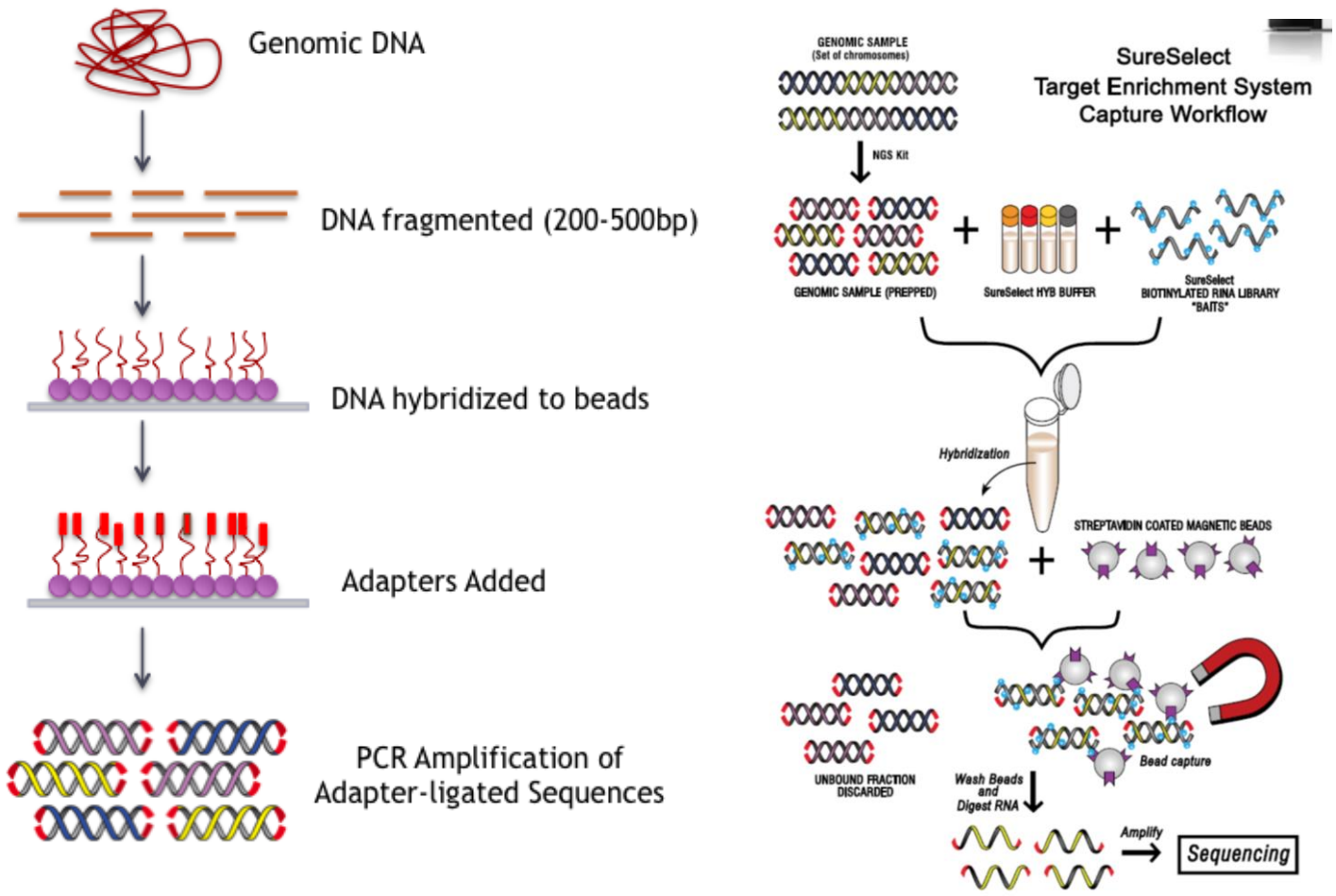


Figure 4. The Exome Sequencing Workflow. *A diagram summarising the processes of library preparation, exome capture, target enrichment and sequencing; (adapted from an image displayed at http://www.genomics.agilent.com).*

1.6b.i Raw Data Processing

In order to process the large amounts of complex raw data generated, there are a series of steps and computer-based tools that must be applied. There are many variations and different combinations in which tools can be used, and therefore, as mentioned previously, only the tools and programs relevant to this study will be discussed.

The raw exome sequencing data is primarily received in the form of FASTQ files; a file format that contains biological sequence. Prior to the alignment to the reference genome, the FASTQ files are analysed using applications such as FASTQC in order to determine various parameters and aspects to the quality of the data. For example, FASTQC enables the user to ascertain the Phred score (an algorithm to measure the quality score of each base call), and also provides information regarding the GC content and read length distribution, in addition to the degree of sequence duplication (Bao et al. 2014). The pre-processed data is then aligned to a reference genome in order to determine regions of variation. In order to ensure a high degree of sensitivity, accuracy and efficiency, there have been many different tools developed to map short reads to the reference genome; Novoalign is one such example, and was the tool used in this study.

After alignment, different computer-based tools such as SAMtools are used in order to convert the large SAMfiles (approximately 20-30Gb) to the more compressed binary BAM file format (approximately 4-10Gb); this enables much more efficient usage of the

data. At this point, the aligned data is then screened and quality filtered to remove duplicates and to realign indels that may have caused sequencing slippage. Variant calling and annotation is then performed using another tool, such as ANNOVAR, which incorporates information from various public databases including dbSNP, 1000genomes, ClinVar, COSMIC and PolyPhen. A summary of the raw data analytical process is outlined in figure 5.

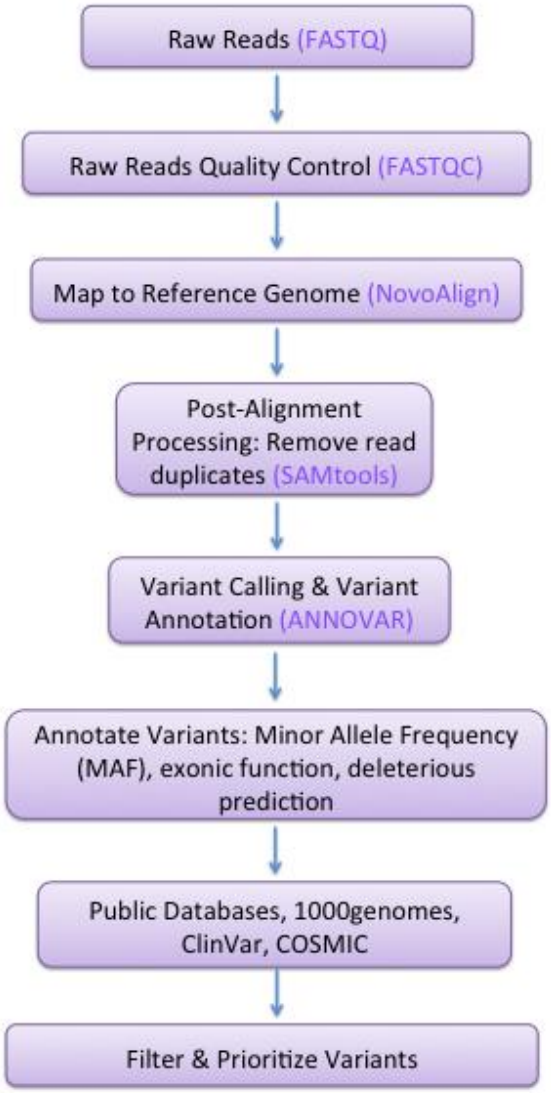


Figure 5. Steps involved in processing raw exome sequencing data. Programs used are written in brackets in purple.

1.6b.ii Sequencing Depth and Breadth of Coverage

As technological advances and abilities continue to progress, the demand for more accessible sequencing at a lower cost per base also grows. Unsurprisingly, sequencing to a greater depth increases the costs, and therefore there have been several studies carried out to assess the minimum amount of coverage required to identify the maximum number of variants; thereby enabling scientists to gain the most accurate and efficient degree of data at the lowest cost.

Coverage refers to the average number of times each nucleotide is sequenced, while the breadth of coverage refers to the percentage of the targeted region of interest that has been sequenced (Sims et al. 2014). During the sequencing process, it is inevitable that technical sequencing errors will be introduced within the data; therefore, for this reason, it is important to sequence a read several times so that the percentage error becomes diluted, and provides a more accurate and reliable result.

When performing exome sequencing to identify disease-associated variants, there is often a delicate balance that needs to be achieved between the degree of coverage, which clearly will improve the accuracy of the variant calls, and the cost per run. This is because the statistical power of disease-gene identification is greatly increased with larger sample sizes; therefore, the lower cost per run allows for a greater number of patients to be sequenced. There have been studies performed with the aim of

ascertaining the ideal limits of this balance. For example, one of the first studies to sequence a human genome using short read methodology was published in 2008; and it found that at a read depth of 15X, nearly all homozygous single nucleotide variations could be identified, and that a read depth of 33X was required to find the same number of heterozygous SNVs (Bentley et al. 2008). It was later determined that a depth of approximately 35x was required to achieve good quality results, and it has since become the standard for the majority of sequencing projects (Sims et al. 2014).

1.6c Applications of Exome Sequencing to Investigate Disease

Next generation sequencing has provided the essential tools necessary to define comprehensive maps and databases of genetic variation including single nucleotide polymorphisms, small insertions and deletions, and structural variants (Cooper & Shendure 2011). Although the majority of these variants are common within the general population, each individual genome will also harbour numerous novel mutations; some of which may be associated with disease. One of the fundamental aims of many human genetic studies is therefore to identify causal variants from a vast quantity of background alterations and to elucidate the roles they play in disease.

1.6c.i Inherited Mendelian Disease and Exome Sequencing

The use of exome sequencing to identify causal variants responsible for Mendelian disease is one of the circumstances in which this technique has had its greatest

successes. This is particularly the case for autosomal recessive disorders, often because all affected individuals will be homozygous carriers of the pathogenic mutation and therefore both parents will be obligate (and most likely heterozygous) carriers of the mutation. For the genetic investigation of Mendelian disorders, large families provide a powerful tool for gene discovery, by providing the opportunity to carry out segregation studies. This technique is made even stronger by the use of consanguineous families, which are known to have more limited gene pools.

This technique has had many successes; one example, Walsh *et al* 2010, used a combination of exome sequencing and homozygosity mapping in a consanguineous Palestinian family, which enabled them to identify the pathogenic mutation (p.Arg127*) in *GPSM2* to be responsible for their recessive deafness (Walsh et al. 2010). Many causal variants continue to be identified in this way, and are proving to be invaluable when providing genetic counselling to both those affected and unaffected carriers. Further identification of disease variants will enable improved pre-conceptive and prenatal screening and counselling, and also may provide information that could lead to novel treatments; at the very least, the identification of previously unreported disease genes can provide affected individuals with a greater understanding of their condition.

1.6c.ii Application of Exome Sequencing in Cancer

Exome sequencing has also made many notable contributions to the genetic investigation of both familial and sporadic cancer. It is more complex to investigate cancers using exome sequencing for several reasons; firstly, the tumour microenvironment is highly genetically heterogeneous, and therefore carries a much larger number of alterations than would be found in constitutional blood DNA. Secondly, most cancer-associated variants have variable penetrance, and so not all carriers of the mutation will have cancer, which means segregation studies are not as powerful as with Mendelian disease. The situation is complicated further by the fact that cancer occurs as a result of both genetic and environmental contributions. Due to the vast numbers of candidate genes received and the limited number of options that can be used to confidently exclude variants, it can be very challenging to successfully identify a candidate gene, and then furthermore, to provide a sufficient amount of evidence that the gene is genuinely involved in increased tumour predisposition or progression. Nevertheless, there have been many achievements in this field which have been due to the abilities of exome sequencing projects. For example, Varela *et al* 2011, used the exome sequencing data of 7 cases of clear-cell Renal cell carcinoma (ccRCC) in addition to the matching normal DNA, and managed to identify a new, major ccRCC gene; *PBRM1* (Varela et al. 2011).

Projects that have been successful have had these achievements largely through the successful identification of variants playing a driver role in initiating or driving tumour progression. By examining the differences between both tumour and normal constitutional blood DNA from an individual, in combination with the DNA of affected individuals with cancer and those without, it is hoped that further cancer-associated alterations may be identified. As this work continues to contribute information, and we continue to learn more about the biological processes involved in the development of cancer, it is hoped that substantial improvements will be made to the predictive, diagnostic, prognostic and therapeutic management of cancer patients.

1.7 Main Aim of Project

The main aim of this study is to employ cutting edge technology to investigate human disease, in the hope that novel disease genes and variants may be identified. Successful identification of disease-associated genes may shed light on new aspects to disease biology and pathogenesis; this information could potentially improve the prevention, screening and diagnosis of these conditions, and could enable the development of novel therapeutic strategies. Exome sequencing has been used to investigate the genetic cause of an autosomal recessive disorder in a consanguineous family, and it has also been used to investigate novel susceptibility genes and genes associated with tumour progression in both familial and sporadic forms of pheochromocytoma and paraganglioma (PCC/PGL).

This study aimed to evaluate the use of exome sequencing as a tool to investigate disease, and through this, demonstrate the future of this technology in modern day genetics. In an ever-approaching era of personalised medicine, it was also hoped that this study could generate potentially useful results that could ultimately be used to improve the lives of those affected by genetic disease.

Chapter Two:
MATERIALS AND METHODS

2.1 Materials

2.1a Patient Material

2.1a.i Acrocallosal Syndrome

Extracted DNA was provided by Dr. Neil Morgan for a consanguineous Israeli-Arab family with Acrocallosal syndrome, and was made up to a concentration of 20ng/µl for each family member.

2.1a.ii Pheochromocytoma

Extracted tumour DNA was received for 31 unrelated patients with sporadic pheochromocytoma, and 25 patients with sporadic paraganglioma. Of these samples, twenty-two were from Japanese patients with sporadic pheochromocytoma, which had been sent by post from Japan by Dr. Masah Yao. Nine of the sporadic pheochromocytoma and five of the sporadic paraganglioma DNA samples were sent from Spain by Dr. Mercedes Robledo, and the final 20 paraganglioma samples were provided by Dr. Tom Martin of the Queen Elizabeth Hospital, Birmingham. Matched normal blood DNA samples were available for all samples apart from the 22 pheochromocytoma patients from Japan.

Anonymised DNA from constitutional blood of 77 unrelated individuals withfamilial pheochromocytomawas obtained from the West Midlands Regional Genetics Lab in Birmingham Women's Hospital.

All DNA samples were received post DNA-extraction at various concentrations. The samples were all diluted to an average and approximate concentration of 50ng/µl and stored at -20°C. The complete clinical history was unavailable for some of these patients. All patients studied gave written research consent and the study was approved by the South Birmingham Research Ethics Committee. Ethical approval was obtained from institutional and local research committees for the collection and use of all samples. Research was conducted according to the principles expressed in the Declaration of Helsinki.

2.1a.iii Control DNA samples

Two 96 well microtitre plates containing approximately 50ng/µl of anonymised control DNA samples from presumed healthy individuals per well were obtained from the West Midlands Regional Genetics Lab. One of the plates contained 96 samples of DNA from Caucasian individuals and the second plate contained 96 samples of DNA from individuals of South Asian ancestry.

2.2 Chemicals, Reagents and Suppliers

Agarose Gel Reagents	Supplier
10 x TBE Buffer (Tris-borate/EDTA)	National Diagnostics
Agarose	Invitrogen
Ethidium Bromide	Sigma
1 kb DNA ladder (1µg/µl)	Bioline
PCR Reagents	**Supplier**
Primers	Sigma Aldrich / Alta Bioscience
Biomix Red	Bioline
GC Rich Solution	Roche
Solutions for Sequencing Reactions	**Supplier**
Ethanol	Fischer Scientific
microCLEAN	Microzone
BigDye Terminator	Applied Biosystems
Big Dye Sequencing Buffer (25mM Tris pH8.7, 4mM $MgCl_{2)}$	Applied Biosystems
Hi-Di Formamide	Applied Biosystems
Whole Genome Amplification	**Supplier**
Repli-g mini Kit	Qiagen
Nuclease-Free Water	Promega
Filter Pipette Tips	Fischer Scientific

2.3 Exome Sequencing

Following informed consent and ethical approval, DNA from 11 patients with familial pheochromocytoma was extracted from peripheral blood by the West Midlands Regional Genetics lab. Approximately 3µg of the extracted DNA was then sent to the Biomedical Research Centre at Guy's Hospital in London for exome sequencing. In addition, approximately 3µg of extracted tumour DNA from 16 patients with sporadic pheochromocytoma was also sent for exome sequencing. In brief, as described by Michael Simpson of the Biomedical Research Centre at Guy's Hospital in London, exon capture was performed using the SureSelect All Exon 50Mb Target Enrichment System from Agilent. Massively parallel DNA sequencing was performed on an Illumina Analyser IIx with 76bp paired end reads. Depth and breadth of sequence coverage was calculated with custom scripts and the BedTools package; a freely available and open-source software (http://bedtools.readthedocs.org/en/latest/). Each run was considered successful and was used for analysis when >80% of the exome was covered to a minimum depth of 20x. Single nucleotide substitutions and small insertion deletions were identified and quality filtered within the SamTools software package and in-house software tools. Variants were annotated with respect to genes and transcripts using the Annovar tool. Filtering of variants for novelty was performed by comparison to dbSNP132 and 1000 Genomes SNP calls (June 2011) and variants identified in 150 control exomes sequenced and analysed by the same method.

2.4 Polymerase Chain Reaction (PCR)

For each candidate gene, primers were designed to flank regions of interest, using Primer3 (http://www.frodo.wi.mit.edu/). Where possible, the set of primers with the smallest difference between the melting temperature of the forward and reverse primers was chosen (supplementary table 1). The primers were then used to amplify the regions of interest from the patient DNA samples.

2.4a Standard PCR amplification for Candidate Genes

All polymerase chain reactions (PCR) were performed on a GeneAmp 9700 thermocycler (Tetrad). Each reaction had a final volume of 25μl, containing between 30ng and 50ng of DNA, 1.5mM Biomix Red and 5pmol forward and reverse primer. Some primers used in the PCR amplifications had a high GC content, and therefore to optimise these reactions, 0.5mM GC Rich Solution was used. To ensure thorough and homogenous distribution of contents, the microtitre plate containing the reactions was placed in a centrifuge and spun at 1000RPM for 30 seconds prior to amplification. For each PCR amplification, an additional reaction was performed using distilled water instead of DNA in order to serve as a negative control. Positive controls were also included which contained a sample of DNA known to amplify well.

2.4b Touchdown PCR amplification

For certain primer sets, a touchdown PCR was required in order to increase the specificity of the region amplified. The reaction was set up as described for the standard PCR reaction. The annealing temperature was initially set to 65°C in order to provide maximum specificity for the primers to bind to the correct region of the DNA. The temperature was then decreased by 1°C every cycle for 5 cycles to maximise specificity of primer binding to produce a large amount of the template target sequence. The annealing temperature then cycled at 55°C to increase the yield further, and in this way help to achieve a higher quantity of specifically amplified DNA than would be present from a standard PCR.

Temperature	Time		Purpose
(a)			
95°C	5 mins		Initial Denaturation
95°C	45 sec		Melting
Ta°C	45 sec	**30 Cycles**	Annealing
72°C	1 min		Extension
72°C	5 mins		Final Extension
(b)			
95°C	5 mins		Initial Denaturation
95°C	45 sec		Melting
65°C *(-1°C every cycle)*	45sec	**10 Cycles**	Annealing
72°C	1 min		Extension
95°C	45 sec		Melting
55°C	45 sec	**25 Cycles**	Annealing
72°C	1 min		Extension
72°C	5 min		Final Extension

Table 1. The Stages of a Standard PCR Cycle. *This table shows the various temperatures involved in a standard PCR reaction (a), and a touchdown PCR (b). The temperature during the annealing stage of the standard PCR was altered appropriately depending on the annealing temperature (Ta) calculated for each of the primers; this was usually between 50°C - 65°C. In the touchdown PCR, the annealing temperature started at a high annealing temperature and was decreased by 1°C every cycle before amplifying at an annealing temperature of 55°C for the remaining cycles.*

2.4c Gradient PCR

In order to optimise certain primer sets, a gradient PCR was performed so as to visualise which annealing temperature yielded the highest concentration of specifically amplified DNA. The reaction was set up as described for the standard PCR. The annealing temperature was altered for each well, ranging from 45°C to 65°C. See *Table 1* for plate set up and temperatures.

2.4d Analysis of Products from Standard PCR Cycle

PCR products were visualised on a 1% (w/v) agarose gel containing Ethidium Bromide. 5μl of PCR product was loaded into the gel, and in addition 5μl of 1kb ladder was also loaded in order to provide an estimate of the size of the bands present. The contents of the gel were then separated by electrophoresis in 1 X TBE buffer at 150v for approximately 30 minutes. The gel was then placed under a UV light transilluminator in order to illuminate the bands and analyse the results.

2.5 Sequencing of PCR Products

2.5a PCR product clean-up

An equal volume of microCLEAN was added to 2.4µl of PCR product in a 96 well microtitre plate, and centrifuged for 40 minutes at 4000RPM. Following this, the supernatant was removed by inverting the plate on to a paper towel, and spinning in the centrifuge at 500RPM for 30 seconds.

2.5b Sequencing Reactions

For each reaction, 2µl of 5 X Big Dye Sequencing buffer was added to 0.5µl of BigDye and 5.5µl of water. Following this, 2pmol/µl of the according forward or reverse primer was added. A PCR reaction was then carried out with the following conditions:

Temperature	Time		Purpose
95°C	30 sec		
50°C	15 sec	**30 Cycles**	Extension
60°C	4 mins		

2.5c Sequencing Reactions Clean-up

The sequencing reaction was then cleaned up by adding 2µl of 0.125M EDTA containing NaOAc to each well along with 30µl of 100% Ethanol, and spinning at 2000 RPM for 20 minutes. The supernatant was then removed again as before, by spinning upside down for 30 seconds at 500 RPM. The pellet was left to air dry for 5 minutes before adding 90µl of 70% Ethanol to each well, after which the plate was spun again at 2000RPM for 10 minutes. The supernatant was then removed again. Finally the pellet was re-suspended in 10µl of Hi-Di Formamide. The microtitre plate was then placed in a GeneAmp 9700 thermocycler and denatured at 95°C for 3 minutes, and finally snap chilled on ice for 5 minutes. Either Dean Gentle or I then loaded the microtitre plate containing the samples on to an ABI Prism 3700 Sequencer. The resulting DNA sequences were analysed using a combination of three different programs; BioEdit Sequence Alignment Editor, ChromasLite and Mutation Surveyor (demo). BioEdit was used to determine an approximation of the number of samples which has been sequenced successfully, while ChromasLite and Mutation Surveyor (demo) were used to visualise the electropherograms and compare them to a reference sequence in order to determine regions of variation.

2.5d Whole Genome Amplification

All DNA samples were amplified using the Repli-G Whole Genome Amplification midi kit from Qiagen. Each kit amplified a maximum of 100 samples with an estimated yield of 40µg. The protocol was followed as per the manufacturer's instructions.

The following buffers were provided in the Qiagen repli-g midi kits and were used in the amplification process: Buffer DLB (contains potassium hydroxide), Buffer D1 (reconstituted buffer DLB and nuclease-free water), and Buffer N1 (stop solution and nuclease free water). In brief, approximately 2.5µl of 30ng of genomic DNA suspended in TE is added to a microcentrifuge tube. An equal volume of buffer D1 was added to each DNA sample and vortexed before centrifuging briefly at 2000RPM. The samples were then incubated at room temperature for 3 minutes. Following incubation, 5µl of buffer N1, containing the stop solution, was added to each sample. The samples were vortexed and centrifuged briefly at 2000RPM.

The DNA polymerase provided with the REPLI-g midi kit was thawed on ice and centrifuged briefly at 2000RPM to ensure homogenous distribution of contents. A master mix was then made up consisting of 10µl of nuclease free water, 29µl of REPLI-g midi reaction buffer and 1µl of REPLI-g midi DNA Polymerase per reaction. To each amplification reaction, 40µ of the master mix was added before incubating at 30ºC for

approximately 16 hours. Following the incubation, the DNA polymerase is then inactivated in each reaction by heating the samples to 65ºC for 3 minutes.

The amplified DNA is then diluted 1:10, and 4µl used in a PCR reaction with a control sample of DNA and a well-known set of primers in order to determine the success of the amplification. Successful amplifications can be identified by running the samples on an agarose gel following PCR amplification, showing a bright band at the expected size.

Chapter Three:

EXOME SEQUENCING AND AUTOSOMAL RECESSIVE DISEASE

3.1 Introduction to Acrocallosal Syndrome

3.1a Consanguinity

Consanguinity is an ancient and deeply rooted societal tradition in certain cultures and can be described as the union of two related individuals who share genes inherited from at least one identifiable common ancestor (Alkuraya 2013). This means that any potential offspring from a consanguineous union will consequentially contain regions of homozygosity throughout portions of their genomes and they will therefore be more susceptible to the manifestation of an autosomal recessive disorder.

In the early years of human existence, close-kin unions were likely to be an inevitable and frequent occurrence as geographical constraints resulted in a limited available gene pool (Bittles & Black 2010). However, in present day, regardless of the freedom to access wider gene pools, in many parts of the world consanguineous unions still persist and they are prevalent across approximately one fifth of the current global population (Hamamy 2012); in fact, it is estimated that at least 10.4% of the world's population have been born to consanguineous parents, where the relation coefficient is limited to second cousins (Alkuraya 2013). Many people perceive that the socio-economic benefits outweigh the genetic health risks associated with a consanguineous union, and therefore the practice of consanguinity still occurs in many populations.

3.1a.i Prevalence of Consanguineous Unions

The degree and extent of consanguinity varies between and within countries throughout the world, but it has been reported to be occurring at its highest rates within certain regions of North Africa, the Middle East and West, Central and South Asia, where close-kin unions collectively account for between 20-50% of all marriages (Teeuw et al. 2013). As figure 6 illustrates, certain populations favour consanguinity, and this can be attributed to several factors including a religious aspect, that clearly has a strong influence on consanguinity, demonstrated by the geographical distribution of its prevalence across the globe (Bittles & Black 2010). The populations in favour of consanguineous unions often gain many socioeconomic benefits from close-kin marriages, which include ease of marriage arrangements, lower divorce rates, lower rates of domestic violence, reduced dowry costs, and the ability to retain finances within the family (Bittles & Black 2010).

For many people, particularly those residing in regions with limited access to education or medical services, the benefits of a consanguineous marriage often far outweigh the genetic risks (Bittles & Black 2010). Furthermore, in these regions, it is often that the main contributing factors to infant and childhood mortality include contraction of infectious diseases and malnutrition; therefore these concerns often overshadow the risks of consanguineous-related autosomal recessive disorders (Bittles 2001; Alkuraya 2013).

However, even in Western countries where the risks of mortality due to infectious disease and malnutrition are rare, some populations, usually minority ethnic groups, favour consanguinity. According to Bittles *et al* 2010, there are approximately 10 million resident migrants in Western Europe who favourclose-kin unions regardless of the common constraints of the Western laws and opinions. Bittles *et al* 2010, reported that there had been a decline in the preference for consanguinity in the Norwegian Pakistani populations recently, possibly due to improved information regarding known health risks; however there had been no reports of decline among the UK Pakistani populations or in Turkish or Moroccan communities residing in Belgium (Bittles & Black 2010).

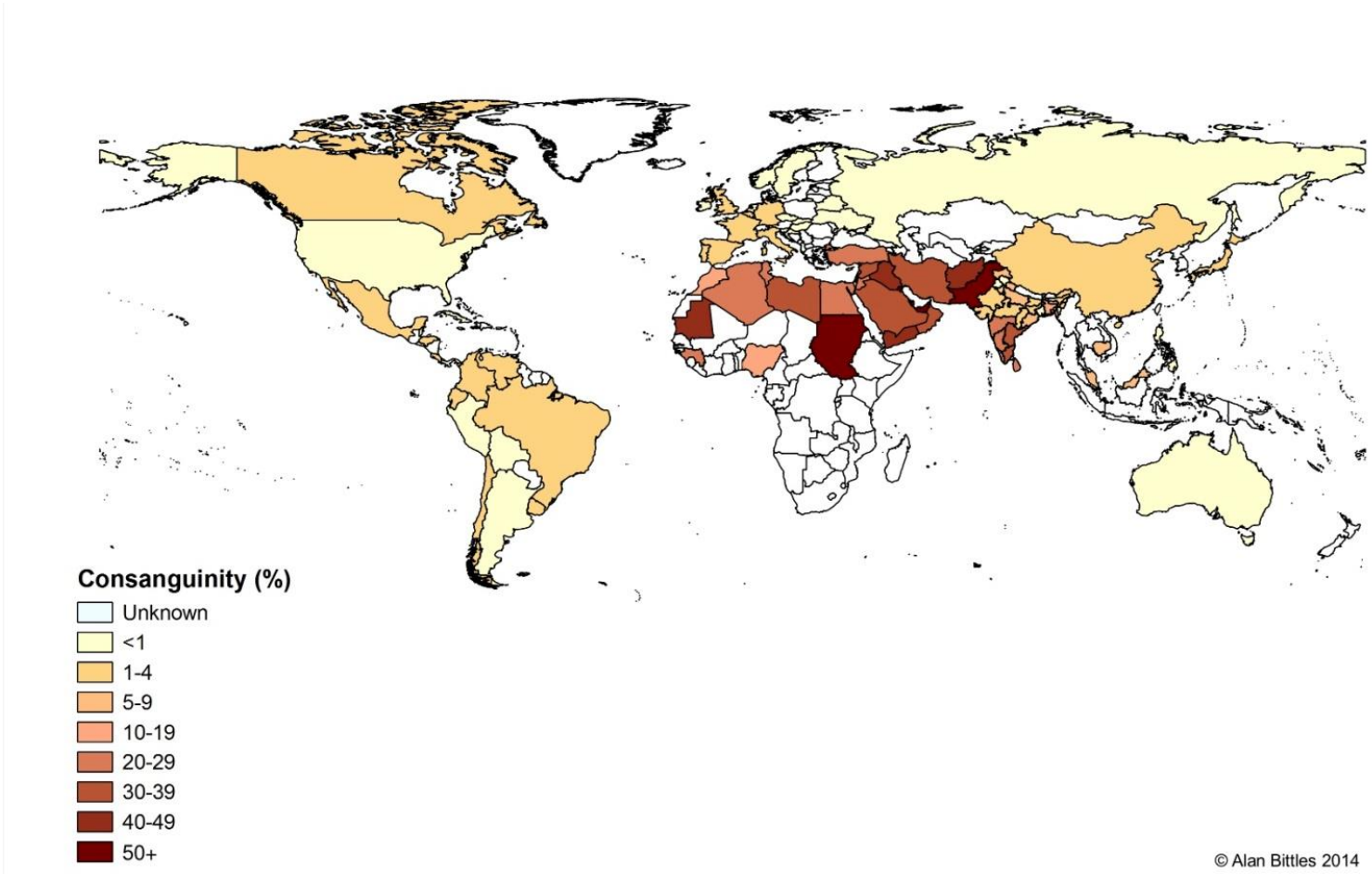


Figure 6. The global prevalence of consanguineous unions (from the global consanguinity/endogamy resource by Bittles et al.2014)

3.1a.ii Genetic Consequences of Consanguinity

The genetic implications to offspring born from a consanguineous marriage are dependent on the proportion of identical gene pairs inherited from a common ancestor, and are therefore also proportionate to the degree of relatedness of the two individuals. For example, first cousins are expected to share approximately 12.5% ($1/8^{th}$) of their gene loci, and so any offspring will be homozygous at 6.25% ($1/16^{th}$) of their gene loci; these genomic regions are then referred to as being identical by descent (Hamamy 2012). This theory is sometimes expressed as the coefficient of consanguinity F) of 0.0625. An example of this can be seen in figure 7.

The genetic risks involved with possessing portions of genome that are identical by descent relate to a higher probability of inheriting two identical and functionally relevant pathogenic alleles which will lead to the increased expression of an autosomal recessive disorder. Studies have shown that birth defects are 2-3% more common in babies born from first cousin relations when compared to other newborns in the general population Hamamy 2012).

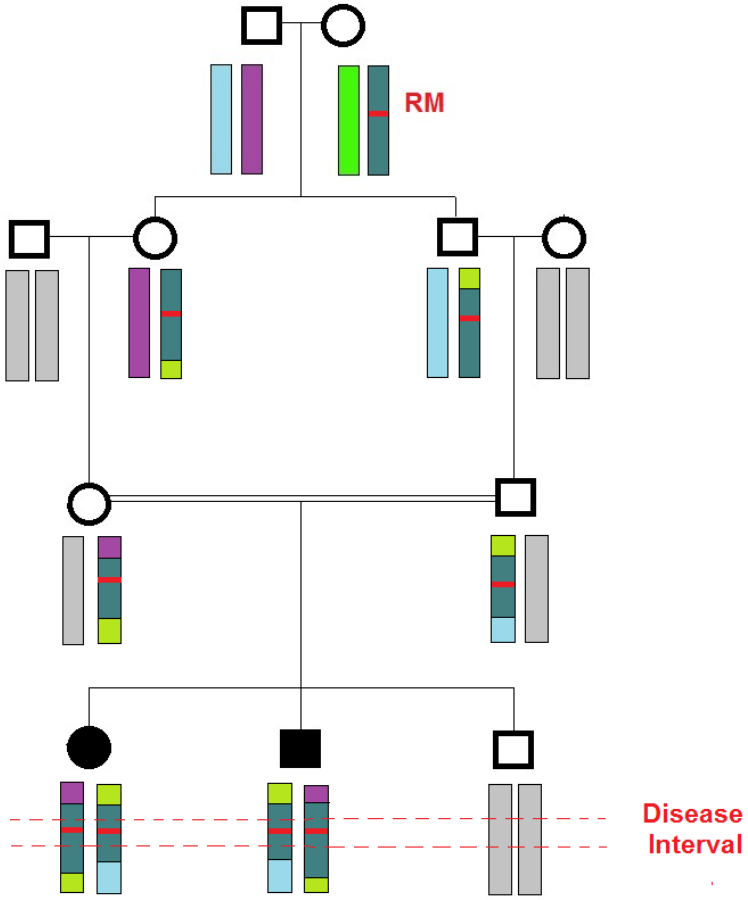


Figure 7. *A diagrammatic example of the potential genetic consequences of a consanguineous union (taken from the MRes thesis of Walsh, 2010). RM refers to a recessive mutation.*

3.2 The Ciliopathies

3.2a The Cilium

The cilium is a multifunctional and evolutionarily conserved organelle that can be found on the apical surface of most vertebrate cells (van Reeuwijk et al. 2011). Functional cilia are known to be essential for a multitude of biological roles, including whole cell locomotion, signal transduction, fluid movement and chemotaxis, among others (Badano, Mitsuma, et al. 2006). Consistent with this fact, genetic mutations resulting in dysfunctional cilia often lead to disorders with a diverse array of overlapping phenotypes, collectively known as the 'Ciliopathies'.

Cilia may exist in a variety of different tubular forms indicative of their tissue-specific functions (van Reeuwijk et al. 2011). Common to all cilia is a ring-shaped scaffolding structure known as an axenome that is enclosed in a sheath referred to as the ciliary membrane (van Reeuwijk et al. 2011). Figure 8 outlines the two different structural formations that cilia are commonly found to have. In a primary, immotile cilium, the axenome contains 9 microtubule doublets and when found in this form, the arrangement is commonly known as the 9+0 configuration; however the structure of a motile cilium, sometimes called a flagella, comprises an axenome housing two central, singular microtubules in addition to the 9 microtubule doublets and this arrangement is referred to as a 9+2 configuration (Badano, Mitsuma, et al. 2006; van Reeuwijk et al. 2011).

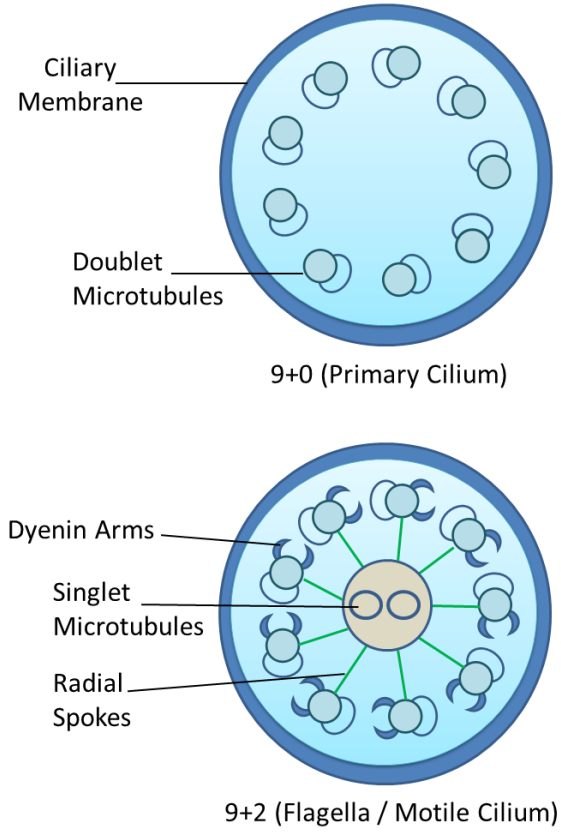


Figure 8. Schematic Representations of the Formations of Primary and Motile Cilia. *This figures shows the structural formations of cilia. The first shows the 9+0 configuration of a primary cilium, and the second image shows the composition of a motile cilium with the 9+2 configuration showing two central singlet microtubule surrounded by 9 doublet microtubules*

Primary, non-motile cilia are essential structures that act as signalling centres for a variety of different cellular signalling cascades, and they can therefore be found in almost every vertebrate cell (van Reeuwijk et al. 2011). The primary cilium has been described previously to act in a similar manner to an antenna, by receiving extracellular

signals through its specialised transmembrane receptors and transducing the signal down through the cilium in order to initiate intracellular signalling pathways which will result in the associated cellular response (Baker & Beales 2009; Singla & Reiter 2006). The essential signalling pathways that are activated through the primary cilium include the Sonic Hedgehog Signalling pathway, the WNT signalling pathway, the FGF signalling pathway, the mTOR signalling pathway, and more recently, the Hippo signalling pathway (van Reeuwijk et al. 2011). It is therefore understandable, given the array of essential signalling pathways communicated through the primary cilia, that mutations resulting in defective cilia-associated proteins can consequentially result in severe disorders; often with a wide variety of different phenotypes. The fact that cilia are known to be so functionally diverse in combination with the fact that research is still uncovering new, previously unreported functional roles could explain why there is such wide phenotypic variability that exists within the Ciliopathies.

3.2b Oligogenic Inheritance and the Ciliopathies

Ciliopathies are most often associated with an autosomal recessive inheritance pattern; however some studies have suggested a more complicated underlying genetic pathogenesis involving oligogenic inheritance. Oligogenic Inheritance refers to the circumstance in which the penetrance or degree of expressivity of a particular phenotypic outcome is governed by the synergistic effects of more than one gene (Badano, Leitch, et al. 2006). Some studies have proposed that the manifestation of

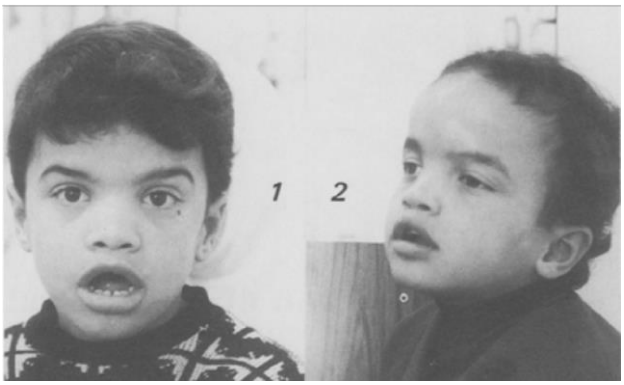
certain ciliopathy disorders is dependent on triallelic inheritance (Katsanis et al. 2001). This was described in 2001, when Katsanis *et al*, genotyped 4 different pedigrees containing Bardet-Biedl syndrome and found that all the affected individuals carried three disease alleles; one homozygous variant in *BBS2* in addition to a heterozygous variant in *BBS6*. Furthermore, the study showed two unaffected individuals in two different pedigrees that carried the homozygous variant in *BBS2*, but lacked the heterozygous variant in *BBS6.* Katsanis *et al* suggested the absence of disease in these individuals was due to the nonexistence of the heterozygous *BBS6* variant and emphasized that it was necessary to inherit all three disease alleles for disease manifestation (Katsanis et al. 2001). Since then, there have been further reports of oligogenic inheritance in other ciliopathies including nephronophthisis, which suggests that this type of inheritance may not be exclusive to Bardet-Biedl syndrome, and may well be occurring in additional, unreported ciliopathies (Hoefele et al. 2007). On the other hand, there have also been a variety of studies that claim to have carried out investigations among large cohorts of Bardet-Biedl patients and found no evidence of triallelic inheritance (Abu-Safieh et al. 2012; Laurier et al. 2006; Hichri et al. 2005; Nakane & Biesecker 2005).

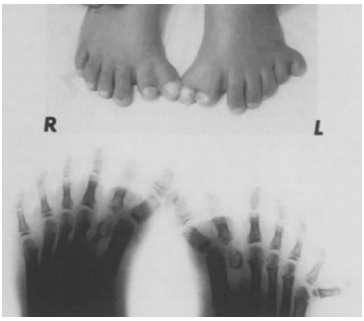
The concept of oligogenic inheritance within the ciliopathy disorders therefore remains a controversial topic. However, given the wide phenotypic spectrum within this group of

disorders and the lack of a more feasible explanation, it remains a possibility that could benefit from investigations in larger cohorts of ciliopathy patients.

3.2c Clinical Features of Acrocallosal Syndrome

Acrocallosal syndrome was first described in 1979 by Albert Schinzel (Schinzel 1979); it is now recognized as a rare and genetically heterogeneous autosomal recessively inherited disorder that is characterised by craniofacial dysmorphism, agenesis or hypoplasia of the corpus callosum and duplication of the phalanges of the hallux/or thumbs (Gelman-Kohan et al. 1991). As displayed in figure 9, common features of Acrocallosal syndrome include craniofacial dysmorphism and polysyndactyly (Hodgson et al. 2009). Acrocallosal syndrome has recently been described to be a ciliopathy disease, due to its overlapping clinical features with Grieg cephalopolysyndactyly syndrome and hydrolethalus syndrome (HLS).





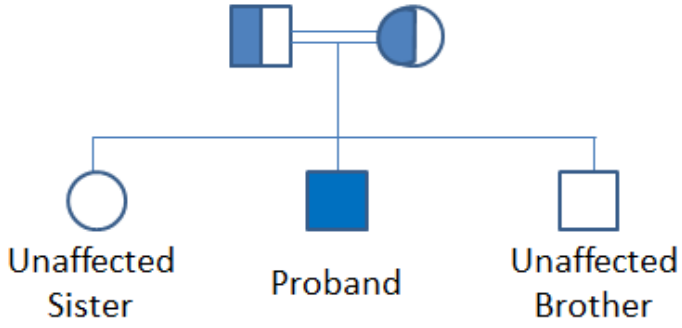


Figure 9. Clinical Features of the Patients with Acrocallosal Syndrome from original report by Gelman-Kohan et al 1991. *The features seen in the first images show evidence of prominent forehead, facial dysmorphism, broad nasal bridge and short upper lip with open mouth. Also shown, is the polydactyly to the left and right feet.A pedigree of the family members for whom DNA was available is also included, displaying the two consanguineous parents, the two unaffected siblings, and the affected proband.*

3.3 Acrocallosal Syndrome and Exome Sequencing: Primary Aims

The primary task of this part of the study was to carry out a 15-year follow up on a previous study carried out on consanguineous family of Israeli-Arab descent comprised of several individuals with Acrocallosal syndrome. The aim of this was to try and identify a causative gene responsible for the acrocallosal syndrome seen in this family. By using next generation sequencing to investigate the exome of an affected individual from this family, it was hoped that the causative variant could be identified, screened in additional family members and confirmed to segregate with the disease.

The proband from the family originally described in the study by Gelman-Kohan *et al* 1991was said to exhibit range of clinical features consistent with acrocallosal syndrome (patient 2 in figure 9). These included macrocephaly, polydactyly of the hands and feet, retardation short philtrum with upturned lip, prominent forehead, and absence of the corpus callosum.

By using next generation sequencing to investigate the DNA from this family, it was hoped that a genetic causecould be identified, which could then provide the family with an identifiable cause for their disease. This information would provide both the family and other affected kindreds access to accurate genetic diagnosis and potential access to pre-marital and pre-conceptive genetic screening and counselling. Furthermore, the identification of a novel variant in a disease gene responsible for acrocallosal syndrome

may shed light on the biology of the disease and potentially provide a novel therapeutic target which could be exploited to provide a treatment.

3.2 Results

3.2a Gene Filtration and Prioritization for Acrocallosal Syndrome

The exome of an affected individual with acrocallosal syndrome was sequenced by next generation sequencing. The raw exome sequencing data was initially analysed at the Biomedical Research Centre at Guy's Hospital; they identified and quality filtered all single nucleotide substitutions and small insertion deletions using a combination of the SamTools software package and their in-house software tools. They then annotated the variants with respect to genes and transcripts using the Annovar tool. They presented all of the data and annotated variants in an excel spreadsheet. At this stage, there were 28,588 variants identified in the proband, and so it was necessary to reduce this number by applying a series of consecutive filters to the data in order to exclude unlikely candidates (this is outlined in figure 10).

The first clear point of filtration was to exclude all variants with a high minor allele frequency (MAF) in the population as it is highly unlikely that a common variant in the general population will result in a rare disorder. The MAF of a variant represents the rarity of the altered allele in the population cohort included in dbSNP (MAF is also

known as GMAF in the population cohort of 1000genomes). It is commonly accepted in many studies that variants portraying a MAF ≥5% are considered common, a MAF between 1-5% are regarded as low frequency variants, and a variant with a MAF <1% is considered to be rare (Wessel et al. 2015; Consortium 2014; Auer & Lettre 2015). The variant annotation, performed by the Biomedical Research Centre at Guy's Hospital as part of the exome sequencing service, included all reported minor allele frequencies from both dbSNP and 1000genomes project (Dec.2010 release).

Given the rarity of acrocallosal syndrome, we initially focused our attention to novel variants only, which reduced the number of candidates to 777. The next filter applied was based on the predicted consequence of the variant type, and was based on the assumption that nonsynonymous, nonsense, frameshift and splice site mutations would be more likely to have a detrimental effect on protein function in comparison to intronic and synonymous variants. Following the removal of all intronic and synonymous variants, the remaining candidates amounted to 332. With the knowledge that recessive disorders are caused by the biallelic inheritance of two mutated alleles in a disease gene, the next filter removed all heterozygous variants, leaving 29 potential candidates.

At this point, the more manageable quantity of 29 variants was investigated further using a candidate gene approach. Genes previously associated with other ciliopathies and those known to play a role in pathways that are commonly aberrant in the ciliopathy

disorders were considered high priority. This process highlighted a nine base pair in-frame deletion (c.653_661delCACGGTCTT, p.(His218_Phe221delinsLeu)) in a gene called *KIF7*. This gene was of particular interest as it codes for a protein that is functional within the hedgehog signalling pathway; a pathway reported to be disrupted in many ciliopathies.

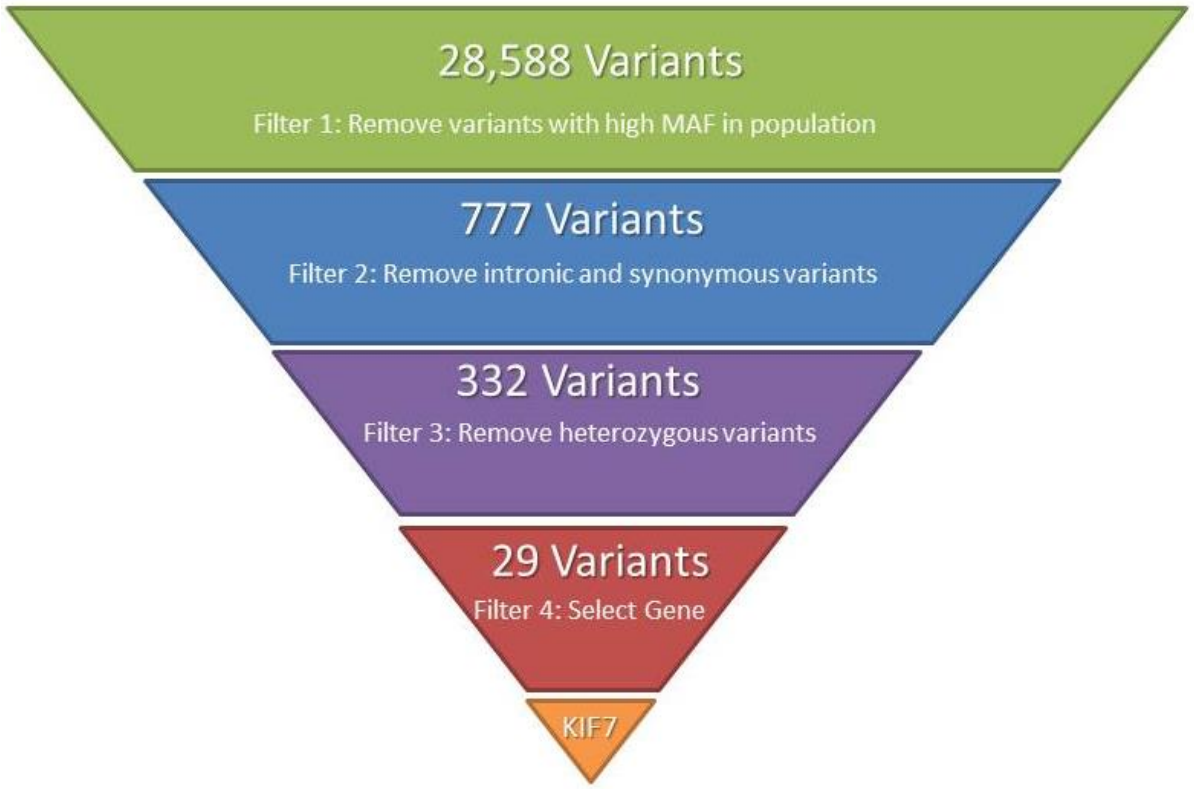


Figure 10. Filters Applied to Exome Sequencing Data for Gene Selection

3.2b Screening of KIF7 in additional family members

In order to determine if the deletion in *KIF7* was likely to be the causative variant, primers were designed to flank either side of the coding exon harbouring the mutation (primers can be found in supplementary table 1).

DNA was available for 4 additional family members, including the mother, father, an unaffected brother and an unaffected sister. Direct sequencing of the PCR amplified region of interest in these samples revealed that the mother and father were heterozygous carriers of the mutation, the proband was confirmed to be homozygous and the two unaffected siblings were both wildtype; the electropherograms of some of the individuals are shown in figure 11.

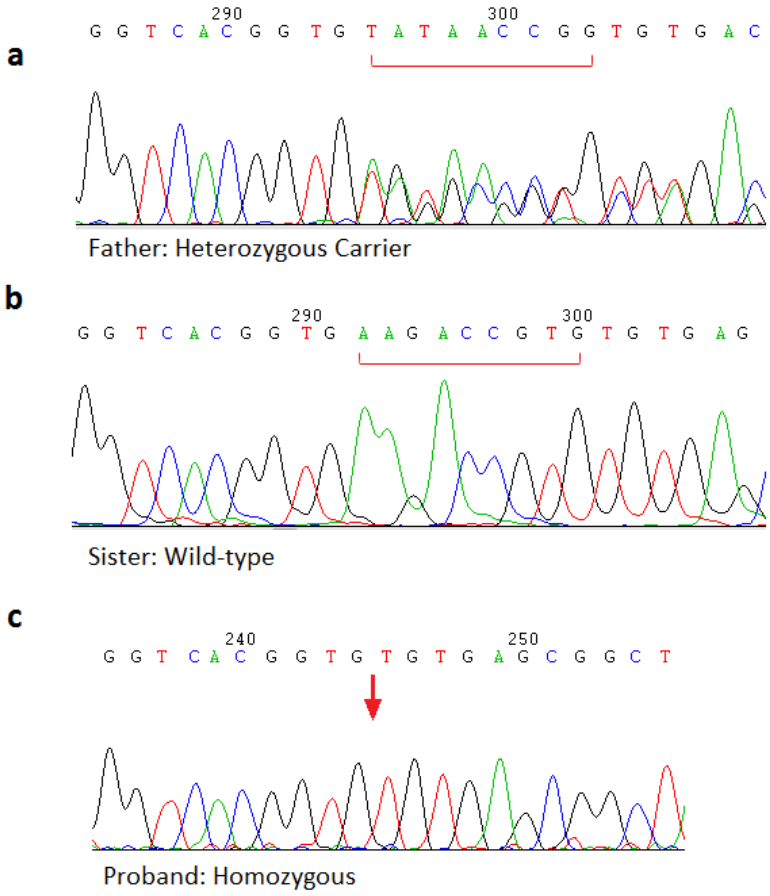


Figure 11. KIF7 Mutation Segregation Status in Family

3.2c Investigation of Evidence for Oligogenic Inheritance

As mentioned previously, some studies have suggested the presence of modifying alleles that are predicted to contribute to phenotypic variability in some ciliopathies (Louie et al. 2010; Katsanis et al. 2001; Badano, Leitch, et al. 2006). Although the concept of oligogenic inheritance within the ciliopathies is a controversial topic, it seemed interesting to investigate the exome sequencing data of the proband with acrocallosal

syndrome in order to formulate our own opinion. Therefore, in order to determine if *KIF7*-associated acrocallosal syndrome might also show evidence of oligogenic inheritance, the exome sequencing data was inspected to identify the presence of previously reported mutations in genes associated with a range of ciliopathies (Joubert, Bardet-Biedl and Meckel-Gruber syndromes). The presence of known ciliopathy-associated variants in the proband, in addition to *KIF7*, would suggest that these variants might be contributing to the phenotypic variability of this disorder.

Interestingly, three heterozygous previously described variants in known ciliopathy genes were detected in *AHI1*, *BBS2* and *BBS4*, all three of which were present in the Human Gene Mutation Database, displayed in table 2.

Gene	Amino Acid Alteration	Minor Allele Frequency (MAF) from dbSNP	PolyPhen Prediction	Associated Ciliopathies
AHI1	p.Arg830Trp	**rs13312995** MAF: 0.0114/57	Possibly Damaging	Nephronophthisis
BBS2	p.Ser70Asn	**rs4784677** MAF: 0.0036/18	Benign	Bardet-Biedl Syndrome
BBS4	p.Met472Val	**rs2277596** MAF: 0.0014/7	Benign	Bardet-Biedl Syndrome

Table 2. Additional Ciliopathy-Associated Variants Identified in KIF7-associated ACLS.

3.3 Discussion

The cause of acrocallosal syndrome in an Israeli-Arab consanguineous family was investigated using exome sequencing. The DNA from one of the affected individuals was sent for exome sequencing and was analysed in an attempt to identify the causative variant. An in-frame deletion in *KIF7* was selected from the list of candidates and, following screening, the deletion was found to segregate with the disease status of the rest of the family.

3.3a KIF7 as a Cause of Acrocallosal Syndrome

Our findings confirm that mutations in *KIF7* can cause acrocallosal syndrome. Accordingly, Putoux *et al* (2011) also described truncating *KIF7* mutations (three nonsense and four frameshift mutations spread throughout the gene) in eight families with acrocallosal syndrome, which supports our discoveries.

KIF7 is a regulator of the evolutionarily conserved Hedgehog Signalling (Hh) pathway; a pathway essential to a variety of embryonic developmental processes. The primary cilium contains receptors that are activated by Hh pathway ligands, such as Sonic Hedgehog (SHH), resulting in a pathway cascade that is transduced down through the cilium tip in order to initiate intracellular responses (He et al. 2014).

The KIF7 protein is 1343 amino acids in length and contains 3 major domains: the motor domain, coiled coil region and cargo domain (figure 12A) He *et al.* 2014, have recently shown that functional KIF7 is essential to the structure of the primary cilium, and is in particular critical for the formation of a specialised compartment at the cilium tip. The principal Hh signalling regulators are highly enriched at the tip of the cilium and are thought to be regulated in the compartment formed by KIF7 (He et al. 2014). It can therefore be inferred that mutations that disrupt KIF7 function will prevent crucial Hh signalling transduction through the cilia as a result of improper compartment formation within the cilium tip. This therefore means that Hh signals cannot initiate the essential

intracellular responses necessary to implement their diverse array of developmental functions. Consistent with these findings, it has been shown that KIF7-deficient mice display phenotypes also found in ciliopathy disorders, including polydactyly, as a result of defective Hh signalling (Ingham & McMahon 2009; Liem et al. 2009).

The in-frame deletion identified in the present study is predicted to disrupt the region flanking the motor domain and the coiled coil region. The amino acids lost as a result of the in-frame deletion (p.(His218_Phe221delinsLeu)) are moderately conserved across multiple species, which indicates that loss of these particular amino acids is likely to affect protein function (figure 12B). This alteration was also absent from dbSNP, and exome variant server, which supports the conclusion that this alteration is not a common polymorphism. Liem *et al* 2009, showed that mutations to *KIF7* affecting the motor domain resulted in decreased localisation of KIF7 to the cilium tip, resulting in loss of KIF7 function and defective hedgehog signalling (Liem et al. 2009). It is therefore possible that the deletion might affect the stability of KIF7, consequentially causing defective hedgehog signalling transduction, ultimately resulting in the phenotypes observed in the affected individuals. In combination with the current literature, our finding therefore supports that *KIF7* mutations cause acrocallosal syndrome in addition to hydrolethalus syndrome and Joubert-syndrome (Dafinger et al. 2011). In addition, our finding also suggests that loss or disruption of KIF7 may contribute to the phenotypes

observed in these disorders through a mechanism that prevents appropriate cellular responses to hedgehog signalling (Putoux et al. 2011; Dafinger et al. 2011).

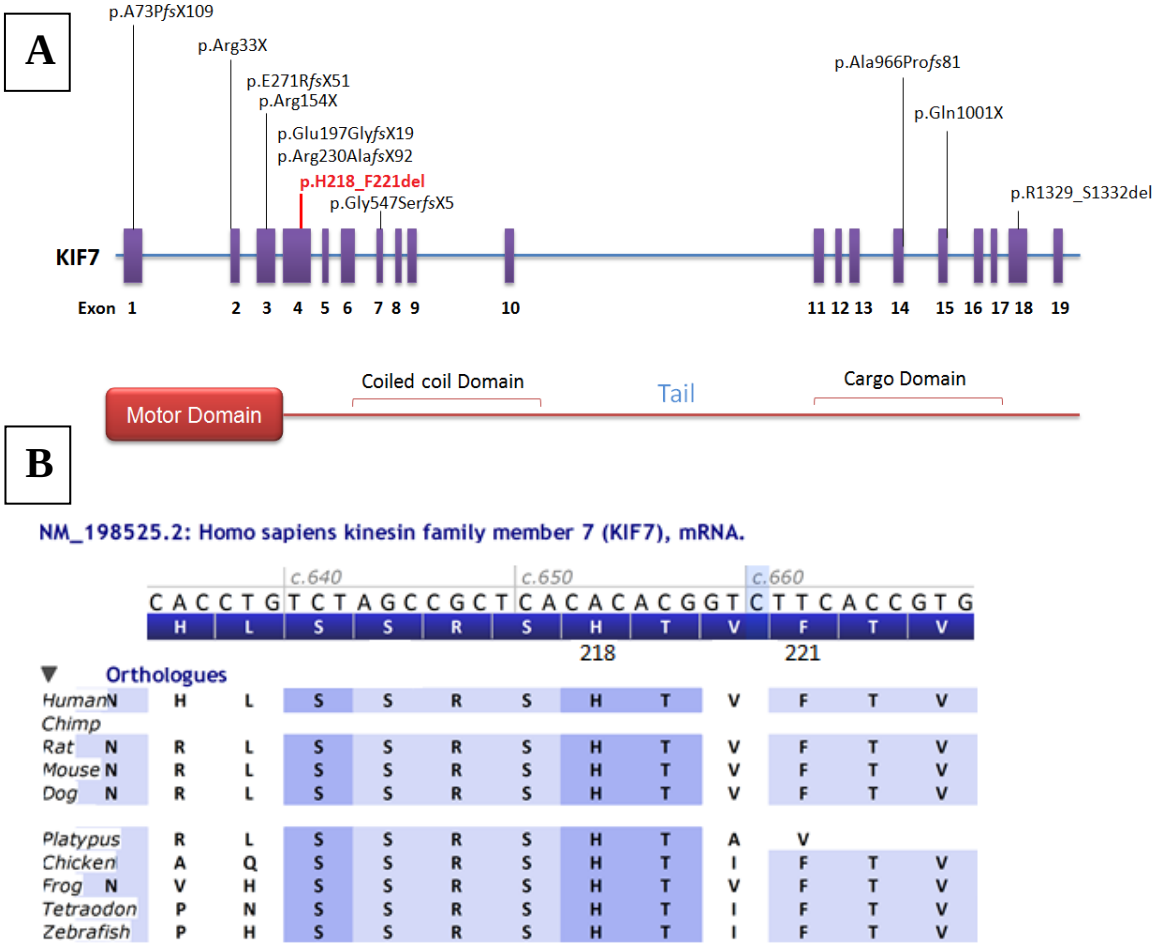


Figure 12. ***A:*** *A Schematic to show the KIF7 gene and the protein domains associated (not to scale), The mutation identified in this study is outlined in red, while previously reported KIF7 mutations associated with Acrocallosal syndrome and Hydrolethalus syndrome are outlined in black.(Taken from Walsh et al. 2013).* ***B****: Shows the conservation across species of the relevant amino acids associated with the in-frame deletion identified in KIF7 p.(His218_Phe221delinsLeu).*

3.3b Oligogenic Inheritance and Acrocallosal Syndrome

The ciliopathies are a diverse and genetically heterogeneous group of conditions comprising at least 29 known disorders, and associated with mutations in at least 89 different genes (McIntyre et al. 2013). The disorders are known to be highly pleiotropic and the penetrance of the known disease-genes can vary between patients. These disorders also exhibit a high degree of variable expressivity, not only between unrelated patients, but even between affected individuals from the same family (McIntyre et al. 2013). In accordance with this, mutations in *KIF7* have been shown to play a role in hydrolethalus syndrome and Joubert syndrome, which exhibit overlapping but distinct clinical features with acrocallosal syndrome including polydactyly, midline brain and facial abnormalities (Dafinger et al. 2011; Putoux et al. 2011).

It is a possibility that the variation in the phenotypes and differences in severities reflect the effects of specific *KIF7* mutations on individual downstream signalling pathways; however it is also conceivable that the phenotypic variation is due to a combined consequence of the functional effects of KIF7 inactivation and the modifying effects of mutations in other ciliopathy genes.

Consistent with this theory, a widening selection of evidence has suggested that primary ciliopathy loci can be modified by the presence of additional mutations in other ciliopathy genes that can exacerbate phenotypes, intensify severities or alter age of

disease onset (Badano, Leitch, et al. 2006; Khanna et al. 2009; Hoefele et al. 2007). Of particular interest was a study by Putoux *et al* 2011 that identified four individuals with Bardet-Biedl syndrome who carried mutations in *KIF7* in addition to mutations in *BBS1*, *BBS7*, *BBS9* and *BBS10*. Putoux *et al* 2011 showed in zebrafish that co-injection of kif7 morpholinos together with the four BBS genes resulted in a higher proportion of embryos with severe phenotypes when compared to injection of kif7 morpholino alone, suggesting an epistatic interaction between *KIF7* and these BBS loci.

Therefore, taking into consideration the current literature, it seemed logical to examine the exome sequencing data in this study to determine whether the potential presence of modifying interactions of other genes was likely to be occurring in the family with acrocallosal syndrome; particularly as a range of phenotypes and severities had previously been reported.

Upon examination of the exome sequencing data of the affected proband, three variants were identified that had previously been associated with ciliopathies; *AHI1:* c.C2488T p.Arg830Trp; *BBS2*: c.G209A p.Ser70Asn and *BBS4* c.A1414G p.Met472Val. The variant in *AHI1* (p.Arg830Trp) was of particular interest, as this variant has previously been reported to have a modifying effect on the phenotypes of nephronophthisis cases; specifically, the mutant T-allele was found to be significantly enriched in cases with nephronophthisis exhibiting retinal degeneration when compared to those cases without

retinal degeneration. This indicates that the presence of the p.Arg830Trp alteration in *AHI1* is interacting with the causative mutations in these nephronophthisis cases to influence a different phenotype (Louie et al. 2010).

The p.Ser70Asn (*BBS2)* and p.Met472Val (*BBS4)* mutations have both previously been associated with oligogenic inheritance in Bardet-Biedl syndrome and a previous study tested the variants for evidence of pathogenicity in zebrafish assays (Zaghloul et al. 2010). Zaghoul *et al* 2010 demonstrated that both mutations were of functional importance based on zebrafish phenotype rescue experiments, in that the phenotypes observed following the administration of the relevant BBS morpholinos could not be rescued by null mutations (including BBS4: p.Met472Val) and in addition, the phenotypes were exacerbated by dominant negative mutations (including BBS2: p.Ser70Asn). These results suggest that the two BBS variants identified in the family with acrocallosal syndrome have been previously shown to be potentially pathogenic, and therefore it is possible that the presence of these variants could be altering or exacerbating the phenotypes observed within the affected family members potentially by interacting with the identified variant in *KIF7*. Furthermore, due to the fact that the p.Arg830Trp variant identified in *AHI1* has previously been reported to have a significant effect on phenotypic variability in nephronophthisis cases by contributing to retinal degeneration, it is possible that this variant might also be interacting with any of the other variants present including *BBS2, BBS4* and *KIF7*.

The contributions of these variants with regard to phenotypic variability in this case of acrocallosal syndrome are at present, unclear; however the evidence present in the current literature strongly suggests that the effects of each variant may act in combination to modify the phenotypic outcome of each patient in this case. This evidence therefore provides a possible explanation to the wide variety of phenotypes and severities among the family members.

Though further studies in a larger cohort of KIF7-associated acrocallosal patients will be required to confirm oligogenic inheritance in acrocallosal syndrome, the hypothesis is that variants in other ciliopathy genes will prove to have an epistatic effect in this disorder, and that the phenotypic variability among the affected individuals of this disorder can be explained by the unique combination of ciliopathy variants they inherit.

Chapter Four:

FAMILIAL PHEOCHROMOCYTOMA AND PARAGANGLIOMA

4.1 What are Pheochromocytomas and Paragangliomas?

Pheochromocytoma and paragangliomas are rare neuroendocrine tumours that arise from the neural-crest derived chromaffin cells of the sympathetic and parasympathetic nervous system (Korpershoek 2011). During embryonic development, the neural crest cells migrate along pre-determined molecular pathways and differentiate into a variety of different cell types including intra-adrenal and extra-adrenal chromaffin cells (Korpershoek 2011; Díaz-Flores et al. 2008). The chromaffin cells within the adrenal medulla receive sympathetic input and can secrete adrenaline and noradrenaline into the blood stream. Subsequently, pheochromocytomas can also be associated with excessive secretion of catecholamines including dopamine, adrenaline, noradrenaline or a combination which can lead to severe hypertension and anxiety if left untreated (Korpershoek 2011).

PCC/PGL tumours display a high degree of heritability, and it is currently thought that approximately 40% of tumours harbour a germline mutation in a known susceptibility gene, either through the inheritance of a pathogenic mutation from a parent, or through the development of a *de novo* germline mutation (Jafri & Maher 2012; Crona et al. 2015; Dahia 2014). There are currently at least 14 genes which have been reported to be associated with the pathogenesis of PCC/PGL tumours, (Astuti et al. 2001; Niemann & Müller 2000; Burnichon et al. 2010; Baysal et al. 2000; Hao et al. 2009; Latif et al.

1993; Zhuang et al. 2012; Wallace et al. 1990; Donis-Keller et al. 1993; Qin et al. 2010; Comino-Méndez et al. 2011; Crona et al. 2013; Clark et al. 2014; Castro-Vega et al. 2014), and it is believed that this number will increase, as some tumours with a strong pattern of heritability do not carry any altering germline mutation in any of these genes.

PCC/PGL tumours are mostly benign and in many cases, can be treated successfully with surgical resection of the tumour; however approximately 10-15% can show metastatic potential (Dahia, 2014). It has been reported that patients with *SDHB* mutations are predicted to have a higher risk of disease recurrence and malignancy, and therefore they are associated with a poorer prognosis (Lepoutre-Lussey et al. 2015); however, apart from this, there are few known genetic indicators of metastatic potential that can be used to predict clinical outcomes in most PCC/PGL tumours.

Many hereditary PCC/PGL tumours form as part of one of several autosomal dominantly inherited syndromes including multiple endocrine neoplasia (MEN2A and MEN2B), Von-Hippel Lindau disease (VHL) and Neurofibromatosis type 1 (NF1). Each of these syndromes is associated with mutations in different genes, which result in a variety of clinical characteristics in addition to the formation of PCC/PGL tumours. For example, patients with NF1 (associated with mutations in the *NF1* gene) can often be recognised by their visually apparent clinical features, including the presence of hyperpigmented spots (known as café-au-lait) on the skin and neurofibromas. Patients

with MEN2A often initially present with medullary thyroid carcinoma, while patients with MEN2B are likely to have additional visible characteristics, such as mucosal neuromas (Maher 2014). Patients with VHL disease are associated with the development of renal cell carcinomas (RCC) and hemangioblastomas. Interestingly, *VHL*-related RCC and hemangioblastomas often harbour highly deleterious and truncating mutations in *VHL*, while PCC/PGL tumours more often contain missense mutations (Dahia 2014).

PCC/PGL tumours normally exhibit one of two transcriptional profiles: cluster 1 contains tumours characterised by a pseudohypoxic signature, and cluster 2 is made up of tumours characterised by an enriched pattern of kinase receptor signalling through hyperactivation of various molecular pathways, including PI3-K, mTOR and MAPK (Jochmanová et al. 2014).

4.1a Cluster 1 Pheochromocytomas and Paragangliomas

Cluster 1 tumours are most often associated with mutations in either *VHL, SDHA, SDHB, SDHC, SDHD, SDHAF2, FH* and *HIF2A*, as disruption to these genes can often generate a pseudohypoxic environment through the stabilisation of Hypoxia-inducible factors (HIFs) (Dahia 2014).

Hypoxia refers to the situation in which the cellular environment does not contain a sufficient concentration of oxygen, thereby requiring it to undergo various adaptive

responses in order to continue and maintain normal cellular processes. A state of pseudohypoxia arises when alterations occur to components of the oxygen-sensing pathway, causing these adaptive responses to occur unnecessarily in the presence of sufficient oxygen concentrations (normoxia). HIFs are heterodimeric transcription factors, made up of an α-subunit and a β-subunit. They function as part of an adaptive cellular response to a hypoxic or pseudohypoxic environment by initiating the transcription of numerous target genes. In the presence of normoxia, PHDs (prolyl-hydroxylase domains) hydroxylate the HIF-α subunits at specific proline residues, which enable HIF to be recognized and targeted for proteasomal degradation by VHL (Marxsen et al. 2004)

4.1a.i VHL and PCC/PGL

VHL encodes pVHL, the recognition component of a tumour suppressor complex with E3-ubiquitin ligase activity, that targets HIF-1α and HIF-2α for proteasomal degradation (Maxwell et al. 1999; Kaelin & Ratcliffe 2008). For this reason, loss of *VHL* causes an accumulation of HIF, consequently resulting in a pseudohypoxic environment and constitutive HIF transcription. This is discussed in more detail in section 5.2c.

4.1a.ii Mutations in TCA-cycle Enzymes and PCC/PGL

HIF stabilisation can also occur through mutation to genes encoding several tricarboxylic acid cycle enzymes. To date, mutations in *SDHA, SDHB, SDHC, SDHD,*

SDHAF2 and *FH* have all been implicated in increased susceptibility to pheochromocytoma and paraganglioma (Burnichon et al. 2010; Astuti et al. 2001; Niemann & Müller 2000; Baysal et al. 2000; Hao et al. 2009; Castro-Vega et al. 2014, Clark et al. 2014). Succinate dehydrogenase is an evolutionarily conserved, heterotetrameric enzyme comprised of four subunits; A, B, C and D. When assembled, the focal role of this complex is to carry out the conversion of succinate to fumarate as part of the TCA cycle. Therefore, mutations to any part of the complex hinders this reaction and causes a consequential accumulation of succinate in the cytosol of the cell (Selak et al. 2005). The role of fumarate hydratase (FH), is then to convert fumarate to malate in the cycle; similarly therefore, loss of function mutations to *FH* also result in an accumulation of both succinate and fumarate (Castro-Vega et al. 2014; Clark et al. 2014). The increasing levels of succinate prevent the association of PHDs with HIFs, thereby preventing HIF hydroxylation, leading to HIF stabilisation (Selak et al. 2005).

FH inactivation has been shown to act in a similar manner to *SDHX* inactivation, in that both succinate and fumarate accumulate, which similarly results in stabilisation of HIF (Clark et al. 2014). This is summarised in figure 13, will be discussed in more detail in section 4.3c.

4.1a.iii HIF2A in PCC/PGL Pathogenesis

It has long been suspected that components of the hypoxia pathway might play a tumourigenic role in PCC/PGL, due to the high frequency of pseudohypoxic transcriptional signatures exhibited in PCC/PGL tumours. However, causative mutations have only recently been identified. The first *HIF2A* mutations found to be associated with PCC/PGL tumours were found in 2 patients who also had polycythaemia (a condition associated with an abnormal increase in the concentrations of haemoglobin in the blood). One of these patients also had somatostatinoma (a malignant pancreatic tumour). Following these reports, a further two patients were subsequently discovered who also had *HIF2A* mutations, PCC/PGL tumours, polycythaemia and somatostatinoma. It was later suggested that these conditions may have a non-coincidental clinical association (Zhuang et al. 2012).

The majority of PCC/PGL-associated mutations in *HIF2A* have been found to be somatic/mosaic, and have been observed to cluster in close proximity to the specific proline residue (531), which is located at a critical HIF-2α hydroxylation site (Maher 2013). Mutations occurring at or within close proximity to this site are likely to alter the conformation of the critical hydroxylation domain, thus enabling HIF-2α to avoid proteasomal degradation. Consistently with this suggestion, *in vitro* studies have shown that *HIF2A* mutations at these specific residues result in reduced associations with

PHDs, reduced VHL binding and a consequential increase in HIF stability. This will be discussed in more detail in sections 5.2c, 5.2g, 5.3a and 5.3e.

HIF-1α and HIF-2α collectively regulate approximately 200 transcriptional targets, and therefore constitutive activation of this pathway by either mutation in *VHL*, *SDHX*, *FH* or *HIF2A* can result in an extensive downstream signalling cascade that can initiate the activation of multiple pathways governing angiogenesis, energy metabolism, cell proliferation and apoptosis (Maher, 2013). Although mutations in these genes explain many PCC/PGL tumours that strongly display characteristics classically associated with cluster 1, a large proportion of tumours that exhibit a pseudohypoxic transcriptional signature do not carry a mutation in any of these genes. This suggests that there are additional components of the hypoxia pathway that are likely to play a role in PCC/PGL development and progression that have yet to be elucidated.

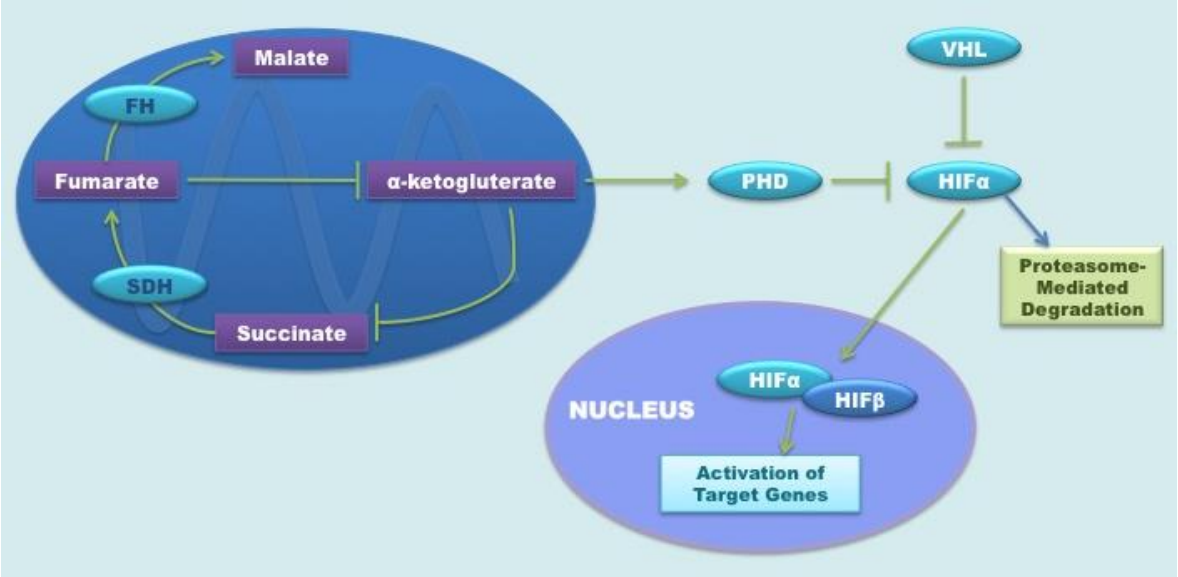


Figure 13. Summary of Cluster 1 PCC/PGL Tumourigenic Pathways; adapted from (Dahia 2014)

4.1b Cluster 2

4.1b.i RET and PCC/PGL

RET encodes a receptor tyrosine kinase protein which plays an essential role during the development of the neural crest; however it also functions to activate various cellular pathways involved in cellular proliferation and neuronal migration. Germline loss-of-function *RET* mutations are a common cause of a condition known as Hirschprung's disease (a condition characterised by intestinal blockages), while gain-of-function mutations are associated with the development of MEN2.

Approximately 95% of the mutations responsible for PCC/PGL development in MEN2A are due to mutations occurring within exons 10 and 11 (Dahia 2014). Mutations in these exons, which make up the cysteine-rich extracellular domain of RET, result in ligand-independent dimerization of RET, which causes its constitutive activation (Bryant 2003). In contrary, mutations in MEN2B nearly all occur within the exonic region encoding the intracellular domain of RET (most commonly, p.Met918Thr) which leads to hyperactivation of downstream targets of RET including mTOR and RAS. The downstream signalling cascade initiated by constitutive RET activation leads to a multitude of pro-tumourigenic cellular consequences including increased signals for growth, proliferation, differentiation and survival. Although *RET* mutations are strongly indicated in MEN2A and MEN2B syndromes, they can also found in up to 5% of sporadic PCC/PGL cases (Dahia 2014).

4.1b.ii NF1 and PCC/PGL

As discussed previously, Neurofibromatosis type 1 (NF1) is a condition that predisposes to multiple tumour types, including among others, PCC/PGL. The *NF1* gene encodes neurofibromin (NF1), a GTPase activating protein that functions as a molecular on/off switch of RAS; NF1 initiates the exchange of the active GTP for inactive GDP, thus terminating RAS signalling.

Loss-of-function *NF1* mutations therefore disrupt the abilities of NF1 to terminate RAS signalling, thus leading to constitutive RAS pathway activation. The tumourigenic effects of RAS pathway activation are discussed in further detail in sections 5.2e and 5.3b; however in brief, constitutive RAS signalling can activate various pathways, including the mTOR pathway, which can then initiate cell growth and proliferation. Consistently, studies in *NF1*-deficient mice show that they develop PCC/PGL tumours with high penetrance (Jacks et al. 1994). It is currently estimated that up to 25% of sporadic tumours carry an *NF1* mutation, which makes it the most frequently mutated gene in PCC/PGL tumours.

4.1b.iii TMEM127 *and PCC/PGL*

The *TMEM127* gene encodes a ubiquitously expressed transmembrane protein that localises to endosomal domains; however, its function currently remains unknown (Qin et al. 2010). Loss of function mutations in *TMEM127* have been discovered to play a role in PCC/PGL tumours, and are thought to exert their tumourigenic effects by causing constitutive mTOR pathway activation. This was originally suggested due to the similarities in the tumourigenic phenotypes between *TMEM127*-deficient tumours and *NF1*-deficient tumours; however the mechanism by which this occurs remains unknown.

4.1b.iv MAX and PCC/PGL

MAX is a protein that is sometimes found as a homodimer but predominantly forms a heterodimer with c-MYC. The MAX-MYC heterodimer has been reported to increase the transcription of more than 1000 target genes which are involved in a variety of cellular processes including cell growth, metabolism and angiogenesis (Blackwood et al. 1992; Dahia 2014). Although the exact mechanism by which *MAX* mutations result in the development of PCC/PGL currently remain elusive, it has been suggested that PCC/PGL tumour susceptibility in the presence of *MAX* mutations is due to the increased transcription of *MYC* target genes which lead to sustained cellular proliferation, growth and increased angiogenesis (Dahia 2014).

As previously mentioned, there are currently at least 14 genes which have been described to play a role in PCC/PGL tumours; however many apparently sporadic tumours display strong hereditary characteristics but do not harbour any altering mutations in any of the currently known genes. This is indicative of the presence of other PCC/PGL associated genes playing a role in these tumours.

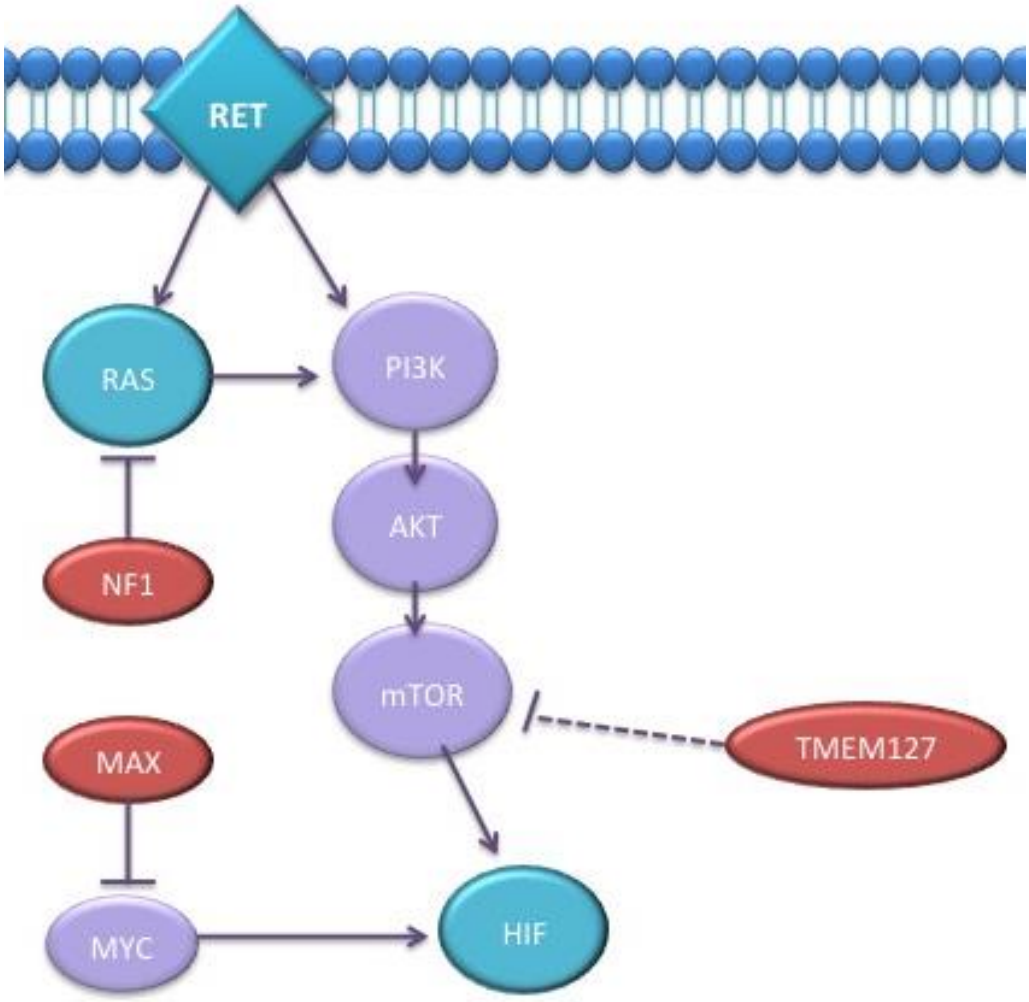


Figure 14 Summary of Cluster 2 PCC/PGL Tumourigenic Pathways; adapted from (Dahia 2014)

4.2 Exome Sequencing & Familial Pheochromocytoma

4.2a Familial Pheochromocytoma and Exome Sequencing: Primary Aims

This part of the study aimed to identify novel causative mutations associated with the inheritance of pheochromocytoma and paraganglioma. This was carried out by investigating the exome sequencing data of 11 affected but unrelated patients who did not carry any known mutation in any known pheochromocytoma or paraganglioma susceptibility gene.

By observing similarities in the exome sequencing data between the 11 affected individuals, it was hoped that a selection of potential candidate genes could be identified and screened in a further cohort of affected individuals. It was hoped that carrying out gene screens in a larger panel of individuals might then reveal the presence of additional variants. Confirming the absence of any identified variants in normal, healthy controls might then indicate that the gene is playing a disease-associated role.

The identification of a novel gene associated with the inheritance of pheochromocytoma and paraganglioma could provide a wealth of knowledge with regard to the biology and pathogenesis of these tumours. This information could then potentially highlight molecular pathways that could be targeted therapeutically, which could create a more

personalised approach to treatments in patients carrying mutations in the associated gene.

4.2b Results

4.2b.i Patient Samples

Eleven unrelated patients with inherited pheochromocytoma were selected for exome sequencing. These samples were selected because they were suspected to have a familial link to the disease, but did not carry any alteration in any of the known susceptibility genes. This combination of circumstances therefore made these patients ideal candidates for the identification of novel, germline variants associated with increased susceptibility to pheochromocytoma and paraganglioma.

Approximately 3 to 5µg of DNA of each sample was sent to the Biomedical Research Centre at Guy's Hospital, where they performed exome sequencing. As before, the group in Guy's hospital identified and quality filtered all single nucleotide substitutions and small insertion deletions using a combination of the SamTools software package and their in-house software tools. They then annotated the variants with respect to genes and transcripts using the Annovar tool, and provided an excel spreadsheet containing all information.

In addition to the 11 samples sent for exome sequencing, a cohort of 77 patients with familial PCC/PGL was used for the mutational screening of candidate genes. The 77 patients in the cohort of familial PCC/PGL samples did not carry any alteration in any known PCC/PGL susceptibility gene.

4.2b.ii Filtering the Exome Sequencing Data

The first step when attempting to analyse this large data set was to combine all of the data from all patient samples into one document for easy comparisons. After combining all of the data sets into a single document, the total number of variants amounted to 263,798. A series of filters were then applied to remove unlikely candidate variants in order to achieve a more manageable selection of genes. The method of filtration is outlined in figure 15.

Initially, all synonymous variants were removed as these changes result in the substitution of an identical amino acid, and therefore they are unlikely to have any effect on protein function.

Inherited forms of pheochromocytoma and paraganglioma are rare in the general population, and so it was expected that variants associated with increased susceptibility to these tumours are also likely to be rare; therefore the next filter removed any variants deemed to be common within the general population. This was done by removing

variants associated with high minor allele frequencies. As mentioned previously, it is commonly accepted that variants portraying a MAF ≥5% are considered common, a MAF between 1-5% are regarded as low frequency variants, and a variant with a MAF <1% is considered to be rare (Wessel et al. 2015; Consortium 2014; Auer & Lettre 2015). Therefore, to ensure only rare variants were kept, any variant reported to have a MAF >1% was excluded.

In addition to removing variants reported to have a high MAF, variants were also removed if they had been found to occur frequently within the in-house data from the Biomedical Research Centre at Guy's Hospital. It was decided that variants that had been previously identified within the in-house data more than 10 times would be excluded. High frequencies of certain variants found in the in-house data may correspond to unreported common polymorphisms that have either not yet been described in dbSNP or 1000genomes or do not have current allele frequencies available. Another possibility is that variants found to occur at a high frequency in-house may relate to common regions of sequencing error, leading to false positive readings.

It is important to note that there is a possibility that some variants found within the in-house data may in fact be more relevant to our study than a simple polymorphism found in the healthy population. This is because samples sequenced with the exome sequencing service are likely to be associated with a particular disorder, and therefore the in-house

variants will have likely been found in people associated with disease; therefore, there is a small possibility that some of these alterations could be pathogenic. It is not known what type of samples have been sequenced with this service previously, and it is therefore possible that other researchers may have submitted pheochromocytoma and paraganglioma samples for exome sequencing, giving rise to the possibility that certain disease-associated variants of interest may be found at a higher frequency in the in-house data than we would expect to find in SNP databases. Therefore, it cannot be completely assumed that variants within the in-house data are due to common polymorphisms or regions of error.

In order to choose an ideal exclusion cut-off for in-house findings, all mentioned variables were considered. As no variant was reported to occur more than 1000 times in-house, it was estimated that 10 findings of a particular variant would correspond to a minor allele frequency of approximately 1%. Therefore, including variants with up to 10 in-house findings was determined to be a suitable limit which would allow the exclusion of as many false positive alterations and unreported common polymorphisms as possible, while ensuring potential variants of interest were retained.

All homozygous variants were also excluded due to the fact that an autosomal dominant pattern of inheritance was expected. Following the application of all of these filters, the number of variants across the 11 patients combined amounted to 3809 variants,

including 2697 missense variants, 566 splicing variants, 67 nonsense variants, 126 frameshift variants and 346 non-frameshift variants.

The next decision was to concentrate primarily on variants that were most likely to affect protein function; so primary focus was given to frameshift, nonsense and splice site variants. Splicing variants were only included where the altered nucleotide was no more than 3 base pairs away from the intron-exon boundary as even though possible, it is unlikely that alterations further than this will cause aberrant splicing. Following the application of these filters, 766 variants remained; 361 of which were classified as novel.

Finally, in a last attempt to reduce the number of candidate variants, the remaining genes were examined in order to identify genes of potential functional relevance by prioritizing those involved in major cancer pathways; particularly kinase signalling, metabolic pathways and hypoxia pathways as these are known to be aberrant in many cases of PCC/PGL.

Genes were also considered interesting if they had previously been associated with related tumour types, particularly kidney tumours, as there is a degree of overlap between kidney cancer and pheochromocytoma susceptibility genes. An additional focus was to concentrate on variants which had a MAF of <0.1%, as it was thought that it

would be most effective to concentrate on variants least commonly identified in the general population.

In addition to the aforementioned selection criteria, certain factors were also considered to be desirable when identifying candidates. One of these desirable criteria was if one gene was common to more than one individual, but not more than 3. This was because, similarities between affected individuals that are absent from the general population might signify a link between the particular gene and the pathogenesis of the disease; however the presence of a gene in more than 3 out of 11 individuals (>27%) is more likely to indicate that the gene is highly subject to variation or that this is a common sequencing error. Another possibility for high numbers of variants in the same gene may be that the gene is very large, and therefore, statistically has a higher chance of acquiring random variation. Genes that are highly variable have a lower probability of contributing cancer susceptibility, as disturbances to protein function may be quite frequent in the general healthy population also. Therefore, when trying to identify PCC/PGL susceptibility candidate genes, those that were present in more than one individual, but not more than 3, were considered interesting.

These final preferential conditions were not set as filters, but acted more as supportive factors when attempting to identify potential candidates. No gene was ever officially excluded from the first round of 3809 filtered variants (outlined in figure 15); however

due to the sheer volume of alterations present in the data, main focus and attention was given to those variants which had passed through the most highly stringent criteria.

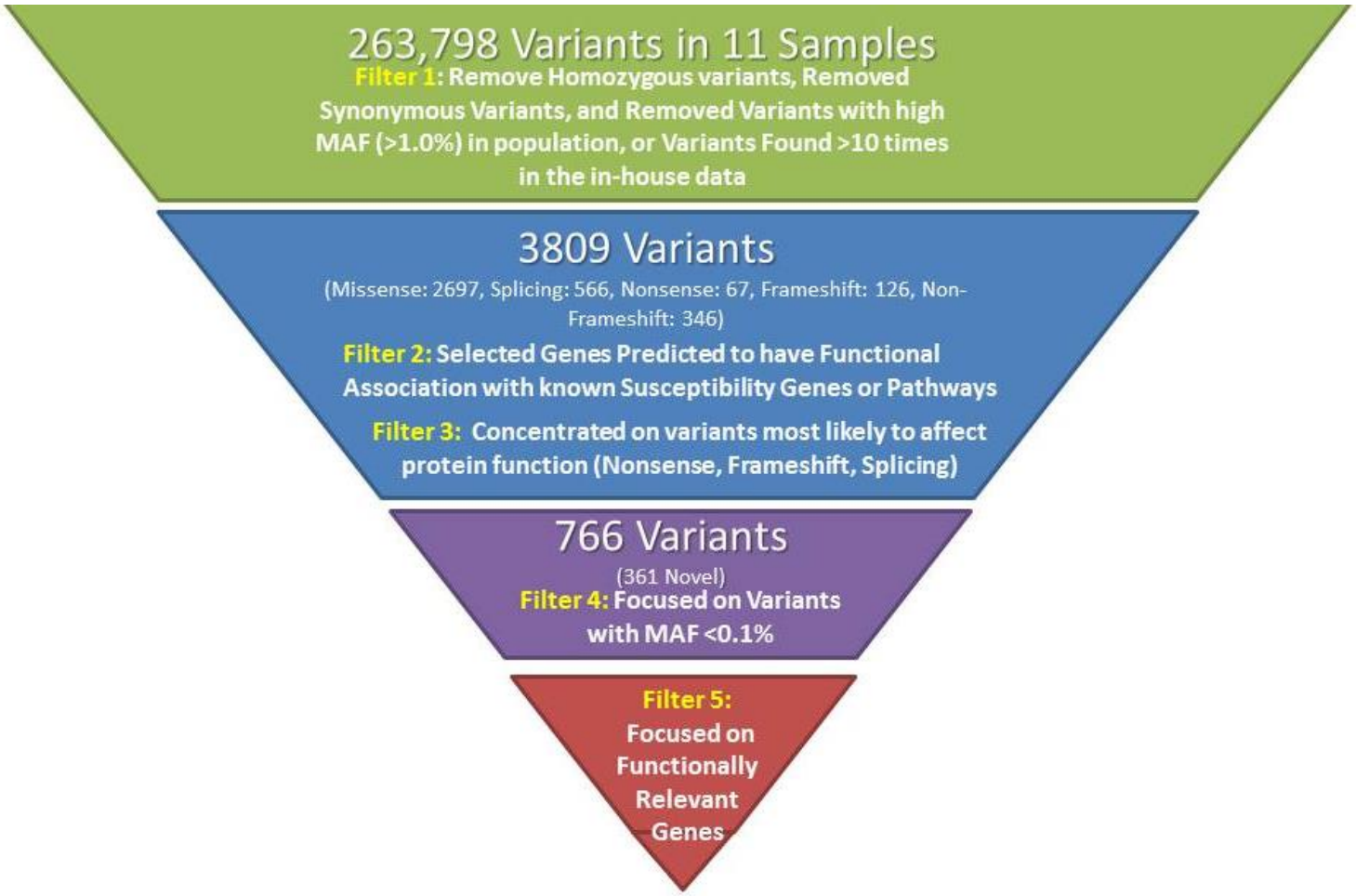


Figure 15. Filtration of Variants for Familial PCC/PGL Exome Sequencing Data

4.2b.iii Candidate Gene Prioritization

A number of genes were identified based on the selection criteria outlined previously; they are summarised in Table 3. Concentration was initially focused on the nonsense mutations, as they are rare and often result in nonsense-medicated decay; thus they are likely to be detrimental to protein function.

The first stage of screening was to confirm the presence of the variant in the corresponding individual in order to exclude the possibility of a false positive. This involved designing specific primers to include the region of interest for each of the candidate variants chosen (primer list found in supplementary table 1). The selected regions were then amplified by PCR in the associated DNA samples, and analysed by direct sequencing.

Patient ID	Gene	Variant Annotation	Variant Type	MAF (dbSNP, 1000g, EVS)	COSMIC	Notes
1F	*ADAMTSL3*	NM_207517:exon18:c.C2230T:p.R744X,	StopGain	Novel	p.R744L in 1 Lung cancer sample (2.4% tumours mutated)	Demonstrated to play important roles in angiogenesis and cell migration. Has previously been found to be frequently mutated in colorectal cancer (GeneCards 2015)
2F	*ARHGEF5*	NM_005435:exon2:c.28_29insGCTGAGGAGGCCCAGCGTGGAG:p.A10_S11delinsGX	StopGain	Novel	-	The encoded protein may form a complex with G proteins and stimulate Rho-dependent signals. (GeneCards 2015)
2F	*CA12*	NM_001218:exon7:c.C718T:p.R240X,	StopGain	0.02% EVS	p.R240Q in 1 Pancreatic Cancer sample	This gene encodes carbonic anhydrase. has been found overexpressed in 10% of clear cell renal carcinomas. (GeneCards, 2015)
3F	*CNTN6*	NM_014461:exon7:c.C728A:p.S243X,	StopGain	-	Gene mutated in 2.2% of tumours in COSMIC	This gene encodes a neuronal membrane protein that functions as a cell adhesion molecule. It may play a role in the developing nervous system. (GeneCards, 2015)

2F	*ERF*	NM_006494:exon4:c.C1603T:p.Q535X,	StopGain	-	-	ERF is a transcriptional repressor, regulated by RAS/ERK pathway. Members of the ETS family of transcription factors, such as ERF, regulate cell proliferation and differentiation. May regulate other genes involved in cellular proliferation. (GeneCards, 2015)
6F	*KIF16B*	NM_024704:exon19:c.C2728T:p.Q910X,	StopGain	0.01% EVS	Q910Q (synonymous)	The protein encoded by this gene is a kinesin-like protein that may be involved in intracellular trafficking. (GeneCards, 2015)
4F	*ME2*	NM_002396:exon8:c.C808T:p.R270X,	StopGain	-	-	This gene encodes a protein involved in the TCA cycle. It is responsible for converting malate to pyruvate. (GeneCards, 2015)
4F	*PCGF5*	NM_032373:exon3:c.T197A:p.L66X,	StopGain	-	-	Component of PRC1-like complex,. Thought to maintain the transcriptionally repressive state of many genes. PRC1 complex acts through chromatin remodelling and modification of histones (GeneCards, 2015)
4F, 5F, 6F,	*TMEM158*	NM_015444:exon1:c.219delG:p.Cys74Alafs*10,	Frameshift	Novel	-	Transcription of this gene is unregulated in response to Ras pathway activation. Proposed to function in a neuronal survival pathway. Transmembrane protein, like TMEM127; identical frameshift in 3 patients. Was found to be false positive, result of sequencing error

						(GeneCards)
5F	*UBE2C*	NM_181801:exon6:c.C412T:p.Q138X,	StopGain	-	-	Ubiquitin conjugating enzyme. Involved in targeting abnormal or short-lived proteins for degradation. Ubiquitination involves at least three classes of enzymes:ubiquitin-activating enzymes, ubiquitin-conjugating enzymes, and ubiquitin-protein ligases. (GeneCards)

Table 3. A Table Outlining the Candidate Variants Identified from Exome Sequencing Data

4.2b.iv Confirming Variants

Primers were designed for each of the variants in order to determine if the alterations were true variants or false positives.

The first filtration process highlighted three identical frameshift deletions (p.Cys74Alafs*10) in *TMEM158* in 3 patients. This was immediately identified as a potential candidate as it passed through all of the stringent criteria and, in addition, it was reported that the transcription of this gene is regulated by RAS pathway activation; a pathway known to be aberrant in many cancers, including PCC/PGL (Weizmann Institute of Science n.d.). In addition, it seemed coincidentally similar to transmembrane protein, *TMEM127,* a known PCC/PGL susceptibility gene (Qin et al. 2010); therefore, although little is known about *TMEM158,* it seemed logical to investigate these three alterations. In order to confirm the mutation, primers were designed to flank the region of interest. The region was then amplified in the three patients by PCR and then sequenced using conventional Sanger sequencing. Upon examination of the results, the three patients displayed a wild-type genotype, which meant that the frameshift (Cys74Alafs*10) was in fact a false positive variant.

All of the nonsense variants were tested in the same manner by designing primers for the regions of interest, amplifying them in the appropriate patients and screening by direct

sequencing. All of the nonsense variants were confirmed in their associated individuals as true variants (summarised in figure 16).

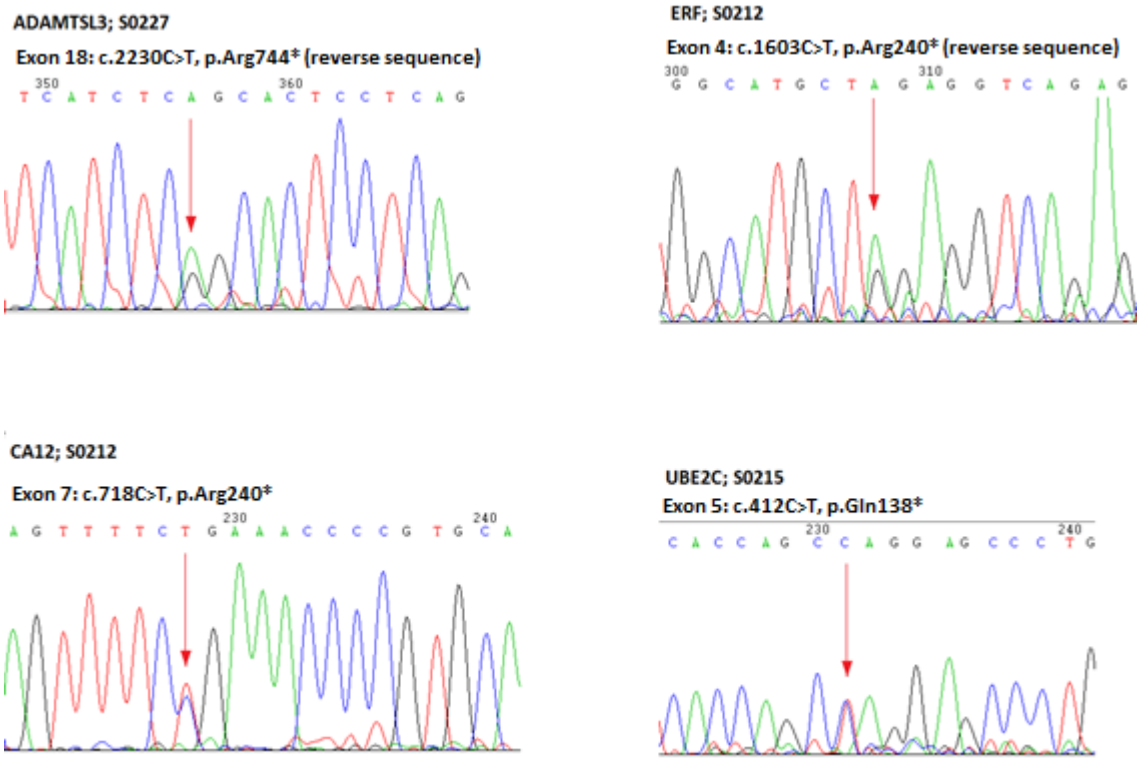


Figure 16. Confirmation of several variants of interest by direct Sanger sequencing.

4.2b.v Screening Genes of Interest in Additional Samples

The confirmation of variants in the associated individuals was the first step to take when attempting to determine if the gene was likely to be playing a role in the development of familial PCC/PGL. Nine variants in nine genes were confirmed to be true alterations. It

was not logical or feasible to carry out full gene screening of all of the 9 genes in the panel of 77 familial DNA samples as this would prove to be a very high workload, and this was not considered to be financially effective. It was therefore necessary to prioritise the genes using a candidate gene approach to determine which genes were likely to be associated with these tumours. Due to the fact that all 9 genes had passed through most of the stringent criteria of the gene filtration process, selecting which genes to screen was largely based on functional relevance and associations with cancer, as reported in the COSMIC database (a database of previously reported, somatically acquired mutations in cancer). After in-depth consideration of each gene, it was decided that *UBE2C*, *ERF* and *ME2* were the best candidates, and so primers were designed to flank each coding exon of these genes and they were then screened in a panel of 77 patients with inherited PCC/PGL.

4.2b.va UBE2C and UBE2QL1

UBE2C is an E2 ubiquitin-conjugating enzyme that plays a critical role in the ubiquitin proteasome system. The ubiquitin proteasome system (UPS) is a physiological mechanism that regulates numerous critical cellular signalling pathways, including cell cycle progression, transcriptional regulation and differentiation (Liu et al. 2015). The ubiquitination process requires the function of three different classes of enzyme; the ubiquitin-activating enzymes (E1), ubiquitin conjugating enzymes (E2), and ubiquitin ligases (E3) (Hershko & Ciechanover 1998). Together, these enzymes orchestrate a

coordinated system to couple ubiquitin molecules to specific protein targets. This enables a variety of cellular functions to take place, including the programmed degradation of cell-cycle regulatory proteins, and the ubiquitin-mediated degradation of various tumour suppressors, oncogenes and signal-transduction system components (Hershko & Ciechanover 1998). Given the vast array of protein targets that affect numerous critical pathways, it is easy to comprehend the oncogenic potential of disruptions to the ubiquitination process. Consistently with this, mutations in *UBE2QL1*, also an E2-conjugating enzyme, had recently been described as a gene associated with renal cell carcinoma (Wake et al. 2013). The fact that there is a large overlap with many of the susceptibility genes for PCC/PGL and renal cell carcinoma, in addition to the functional relevance of the ubiquitination process in tumourigenesis, made *UBE2C* an interesting candidate to pursue. Although mutations in *UBE2QL1* had not been identified in the exome sequencing data, this was also selected as candidate due to its similarities to *UBE2C*, and it's associations with RCC.

Primers were designed for all 6 coding exons of the *UBE2C* gene, and were used to screen 77 familial PCC/PGL samples for mutations. In addition, this gene was also screened in a Spanish cohort of familial PCC/PGL samples by Dr. Mercedes Robledo *et al* as part of a collaboration*;* and therefore in total, *UBE2C* was screened in a cohort of 316 PCC/PGL patients. A summary of the mutation screening is outlined in table 4.

Exon	Primers	Conditions	Variants	Frequency
1	**F:**GCCAGTGGGTAGGTCTAGCA **R:**GAGCTCCTGGTGTGTTCTCC	T.D (65°C - 55°C)	1. p.Ala10Ser 2. p.Ser23Arg	1. 1/316 2. 5/316
2	**F:**ATTCAGGTGGGAGAAGATGC **R:**TGGTGTAGGTGGGGTGTATG	T.D (65°C - 55°C)	c.129+109delC (DIV)	-
3	**F:**GAAAGCTCACCCACTGACCT **R:**TGCACAGAGGGAAAGAACAA	T.D (65°C - 55°C)	No Variants	-
4	**F:**ATGAGCCCAGGAGTTTGAGA **R:**CAGGGAGTCCTGACAGGTTC	T.D (65°C - 55°C)	No Variants	-
5	**F:**GTGGACACCCAGGGTAACAT **R:**CTGGGCTCCTTTTGTTTGAG	T.D (65°C - 55°C)	No Variants	-
6	**F:**CCAGCTTGGTCAACAGAGT **R:**GCCTACCCTGAGTACCTCC	T.D (65°C - 55°C)	No Variants	-

Table 4. Variants Identified in UBE2C (NM:181801, NP:861517) from 316 familial PCC/PGL patients. *Variants of unknown significance are outlined in green and previously reported polymorphisms are outlined in purple. "DIV" refers to a deep intronic variant. Where possible, the frequency of the identified variant from the cohort is included. Further information on the previously reported polymorphisms can be found in supplementary table 2.*

The mutation screening results for *UBE2C* revealed one variant of unknown significance, one known polymorphism and one deep intronic variant from 316 patients. *In silico* investigations using PolyPhen (a web-based tool that predicts the functional impact of mutations) suggested that the missense alteration (p.Ala10Ser) was likely to be benign (figure 17).

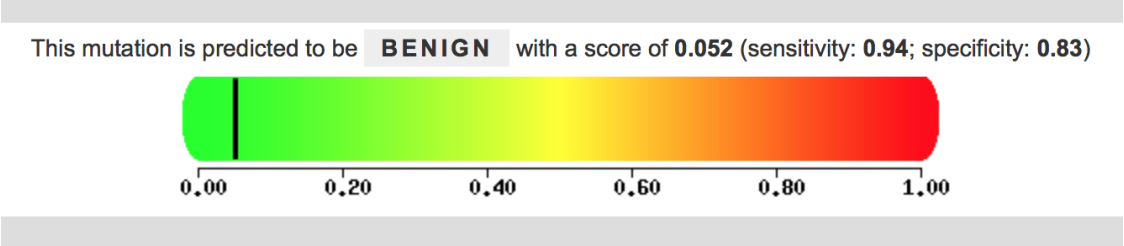


Figure 17. PolyPhen Prediction of p.Ala10Ser Variant Identified in UBE2C. *This in silico tool predicted that the substitution of Alanine for Serine at codon position 10 in the UBE2C gene would have a benign effect on protein function.*

Primers for all coding exons of *UBE2QL1* and their reaction conditions were provided by Dr. Naomi Wake. Mutation screening was then performed in a cohort of 38 familial PCC/PGL samples. Mutation screening did not reveal any variants of significance (data not shown).

4.2b.vb ETS2 Repressor Factor (ERF)

ERF is part of the ETS family of transcription factors; one of the largest protein families of transcriptional regulators. There have been over 200 target genes associated with the ETS transcription factors, collectively making them responsible for the regulation of numerous cellular processes including cellular proliferation, differentiation, development, hematopoiesis, apoptosis, metastasis, tissue remodelling, angiogenesis and transformation (Sementchenko & Watson 2000). The majority of the ETS family of proteins function to activate transcription; however, ERF acts as a transcriptional repressor with tumour suppressor abilities. Studies have shown that the tumour suppressor activities of ERF are highly regulated by the RAS/ERK signalling pathway; a pathway which is commonly activated in RAS-associated pheochromocytomas. It was therefore hypothesised that the nonsense alteration (p.Gln535*) identified in the exome sequencing data in *ERF* may potentially disrupt ERF function, therefore causing a consequential reduction or loss of its tumour suppressive protection. The fact that activation of the RAS/ERK pathway has been shown to promote tumour progression in certain pheochromocytomas suggests that ERF, a downstream target of the same pathway, may also have a role in these tumours.

Primers were designed to flank all coding exons of the *ERF* gene. The fourth exon was divided into four overlapping separate primer sets due to its large size. All coding regions of *ERF* were PCR amplified and sequenced by direct Sanger sequencing in order

to determine if any further loss of function mutations could be identified in a cohort of PCC/PGL patients. The primers, their conditions and the sequencing results are outlined in table 5.

ERF Exon	Primers	Conditions	Variants	Frequency	PolyPhen Prediction
1	**F**:CCCGGAGTGCGATATTAACC **R**:CCACGAGAAGCAACAGGTCT	With GC-Rich, 58°C, 35 cycles	No Variants	-	-
2	**F:**TTGACTTGGCTGAGGTGAGA **R:**TCAGACCCAATGCTTGTTCC	T.D (65°C - 55°C)	1. p.Glu19Val (false positives)	1. 7/77	1. Prob Dam.
3	**F:**ACGTACCTGACCTTCCCAAT **R:**GGAAGATGAAGATGAAGAGCAG	T.D (65°C - 55°C)	1. p.Leu81Leu	1. 2/77	-
4(i)	**F:**CTACATGCCCCTTCAGGCTA **R:**GGGGATAGACTCGGAAGACA	T.D (65°C - 55°C)	1. c.373 -1 G>A 2. p.Arg205Arg	1. 1/77 2. 5/77	1. Benign
4(ii)	**F:**TAGTGATGGCACGTCAGAGC **R:**TTGTAGACGCTTTGGGTGTG	T.D (65°C - 55°C)	1. p.Arg205Arg	1. 5/77	-
4(iii)	**F:**AGGCTCCCACTTCTCCTTCA	T.D (65°C - 55°C)	1. p.Ala410Ala 2. p.Ser327Ser	1. 3/77 2. 6/77	-

	R:CCGTCTTCCTCATC CTCATC				
4(iv)	**F:**CACCACAGATC AAGGTGGAG **R:**GGCAGGGAAG AGACAAGAGA	T.D (65°C - 60°C)	1. p.Glu470Val	1. 26/77 (also found in healthy controls)	1. Poss Dam

Table 5. Summary of Variants Identified in ERF. *ERF (NM_006494, NP_006485) was sequenced in 77 familial PCC/PGL patients. The primers used and their conditions are included in columns 2 and 3 (TD refers to touchdown PCR reactions). Previously reported, known polymorphisms are included in purple, novel, synonymous variants are noted in orange and variants of unknown significance are outlined in green. Where possible, the frequency of variants identified in the cohort and their predicted functional consequence is also outlined.*

The mutation screening process was carried out using whole genome amplified DNA. Any queries were repeated using both the whole genome amplified DNA and the original stock DNA samples in order to exclude the possibility of false positives that may have been acquired during the whole genome amplification process.

The sequencing results initially revealed 7 identical missense alterations in exon 2 (p.Glu19Val), and 26 identical missense alterations in exon 4iv (p.Glu470Val*). In silico* investigations showed that these variants were absent from SNP databases including dbSNP and the exome variant server; in addition, both p.Glu19Val and p.Glu470Val were predicted to be probably damaging and possibly damaging respectively by

PolyPhen. PCR amplification and Sanger sequencing reactions were repeated in both the whole genome amplified DNA and the original stock DNA samples. These variants were confirmed in both whole genome amplified DNA samples and the original stock samples. These results confirmed that 9.1% of the PCC/PGL cohort tested positive for the p.Glu19Val alteration and 33.8% tested positive for the p.Glu470Val alteration.

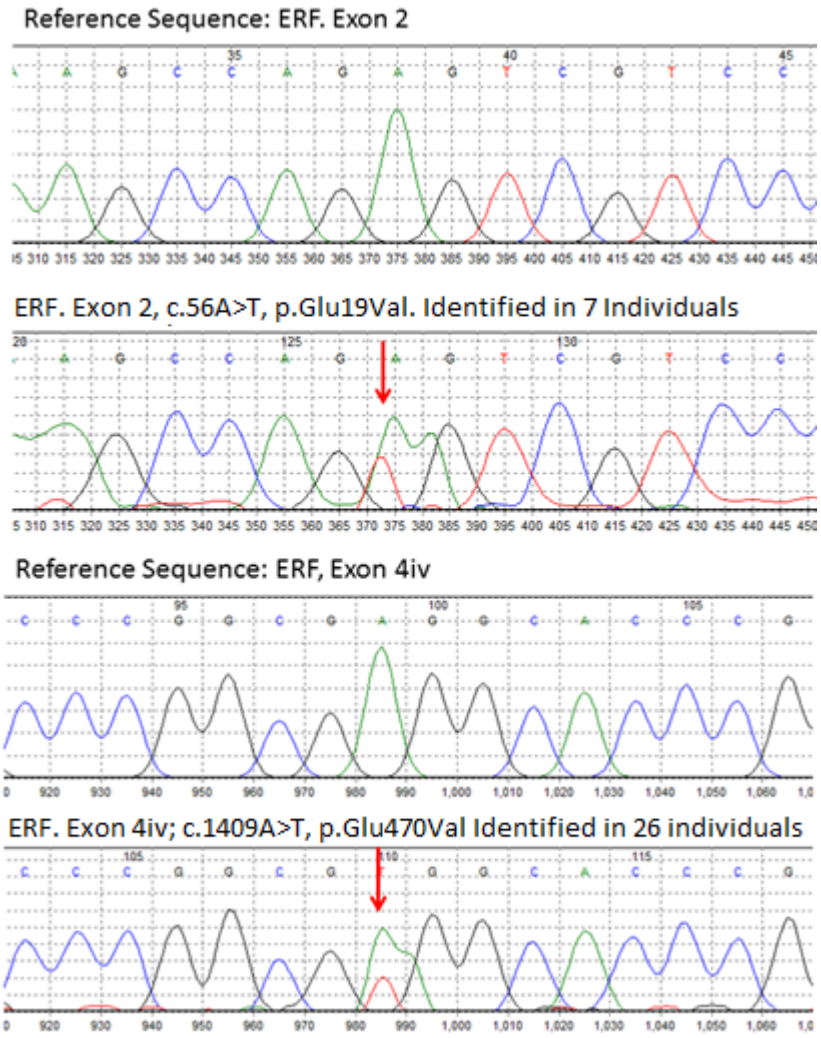


Figure 18. Variants in ERF: p.Glu19Val in exon 2- Identified in 7 individuals and p.Glu470Val in exon 4iv- Identified in 26 individuals.

Although these variants were absent from SNP databases, it was necessary to examine a large set of healthy, ethnically matched control samples in order to explore the possibility that these variants may be the result of unreported common polymorphisms. A cohort of 192 healthy, ethnically matched control chromosomes was screened for the

p.Glu19Val and p.Glu470Val alterations. The results of the sequencing showed that both variants were found to be present within the healthy control samples, which suggested that p.Glu19Val and p.Glu470Val were likely to be unreported, common polymorphisms. The presence of these variants in healthy controls was a strong indication that these alterations were unlikely to be associated with disease.

In addition to these variations, one splicing alteration (c.373 -1>A) was identified, 1 nucleotide upstream from exon 4 (shown in figure 19). This was particularly interesting, as an alteration within such close proximity to the intron-exon boundary could potentially disrupt the splicing within this region, possibly resulting in a truncated protein that could disrupt ERF function.

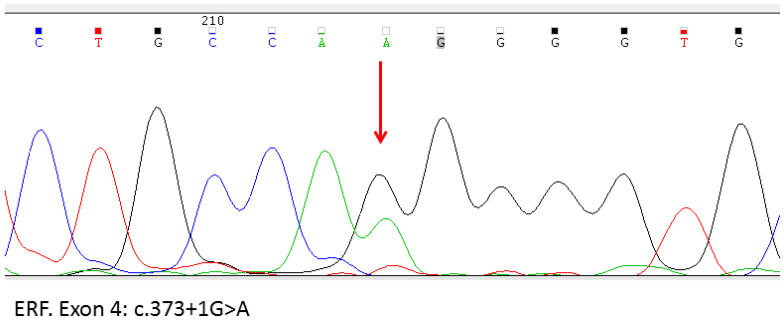


Figure 19. Electropherogram Showing a Heterozygous Splicing Alteration in ERF (c.373+1G>A).

The PCR amplification and direct Sanger sequencing reaction was primarily repeated using the whole genome amplified sample which confirmed that this alteration was not an artefact, and that the variant was genuinely present within this sample. However, to determine if this alteration was truly present in the individual (and not an artefact that had been acquired during the WGA process), repeats were carried out using the original stock DNA. Multiple repeats in the original stock DNA revealed that although the variant was present in the WGA DNA sample, the variant was not actually present in the original stock DNA sample.

In addition to the aforementioned variants detected, several synonymous variants were also identified; however, as none of these synonymous changes were located within close proximity to any intron-exon boundaries, it was unlikely that they will have had any observable effects on protein function. Several common SNPs were also identified

within the cohort, as can be expected; however, no further variants of significance were found to be present in the cohort of 77 familial PCC/PGL patients.

4.2b.vc ME2

Primarily, *ME2* was selected as a candidate gene because the information in The Cancer Genome Atlas (TCGA) showed that many tumours had been found to harbour deletions and mutations in *ME2* as outlined in figure 20. However, the most persuasive aspect of this gene was that it is known to be involved in the tricarboxylic acid cycle (TCA-cycle). ME2 catalyses the oxidative decarboxylation of malate into pyruvate (discussed further in section 4.3c. It is well known that mutations in genes acting within the TCA-cycle can be associated with increased pheochromocytoma and paraganglioma susceptibility, including mutations in *SDHA*, *SDHB*, *SDHC*, *SDHD*, *SDHAF2* and *FH*. It is reasonable to expect that disruptions to different parts of the same pathway might generate similar biological consequences; it was therefore hypothesized that alterations to ME2 function may result in deregulation of the TCA-cycle, thereby yielding similar outcomes as mutations in *SDHx* or *FH* with regard to tumour predisposition.

Primers were designed for all coding exons of *ME2* in order to amplify the selected regions by PCR reaction in a cohort of familial PCC/PGL patients. After PCR amplification of these regions, were screened for mutations by direct Sanger sequencing.

Mutation screening was performed in a panel of 77 whole genome amplified familial PCC/PGL samples. As artificial variants can sometimes be acquired during the whole genome amplification process, any queried variants were repeated in original stock DNA samples to ensure accurate results.

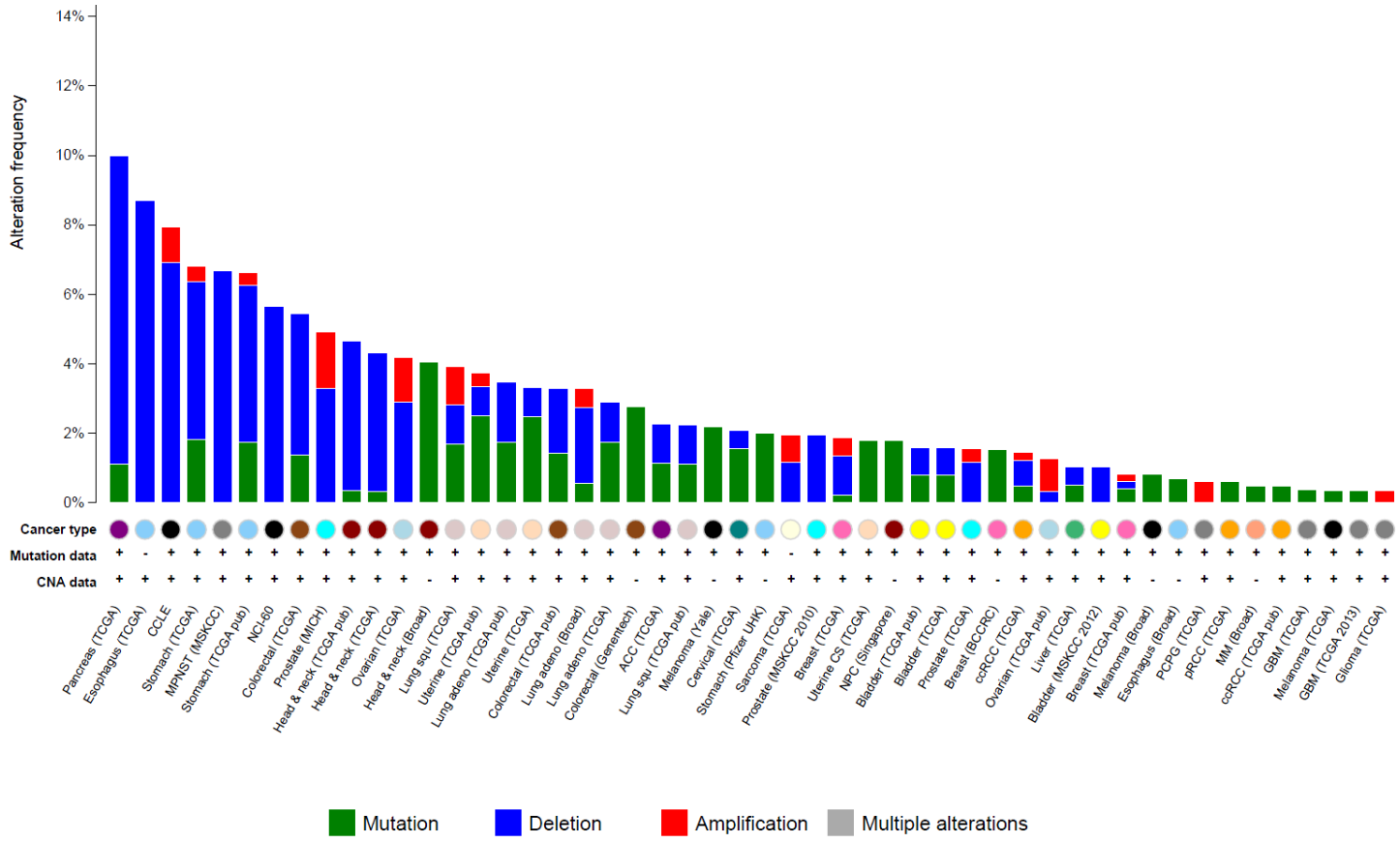


Figure 20. Cancer Alteration Summary taken from The Cancer Genome Atlas (89 Studies for ME2) *This figure displays the high frequencies of deletions and mutations in ME2, in a variety of cancer samples.*

Exon	Primers	Conditions	Variants
1	Non-Coding	-	-
2	**F:** ATACCCGATGGACATGAAGG **R:**TCTGAACAACCTTAGCATAAAGC	T.D (65°C - 60°C)	None
3	**F:**TGTCATTGCTACCCCAGTAAG **R:**CACCGAAGTGAAAGAAACACAA	T.D (65°C - 60°C)	None
4	**F:**GATGTGAGAATGGGGCTCAA **R:**AGGTGGGAAATGTGATTTGG	T.D (65°C - 60°C)	None
5	**F:**GCTTGGAAATTATACCATGTCTTG **R:**TGAGAGCAACAATGAAAAACTG	T.D (65°C - 60°C)	None
6	**F:**GTGCCCCATTTTTGACAACT **R:**ACAGGTATGAGCCACCATGC	T.D (65°C - 60°C)	None
7	**F:**AGCCTGGGTGACAAGAGTGA **R:**TCGCCTCCTGCTTATCTACA	T.D (65°C - 60°C)	None
8	**F:**GGGAATATTAGGTGGAAAGCTG **R:**GCAGCTGTCCCTGTAAGAAAAA	T.D (65°C - 60°C)	None
9	**F:**CAACTATTGTGGGTTTTGCTG **R:**TGAAAGAGCATAGTGAGCCTG	T.D (65°C - 60°C)	None
10	**F:**TCAAGCATTGCATCGTACTC **R:**CTGTTGAAAAGAGAGTGGAAGG	T.D (65°C - 60°C)	None
11	**F:**GGAAGGTATGGGTTGAAGAGG **R:**CATGGTAATGATGACTGGCAAC	T.D (65°C - 60°C)	None
12	**F:**TTGCATTGCTGAGATAGTGC **R:**CATGCTGTAATCTCTAGGGAAACC	T.D (65°C - 60°C)	None
13	**F:**CACCTCAATTTTCTCATCCACA **R:**TCCCATTCCAAAGAATTCCA	Std. 58°C	p.Gly450Glu (10/55)

14	**F:**TTGCGATGTGGATGGATTAC **R:**CCCATTCTCTCTCTGCCTTT	T.D (65°C - 60°C)	None
15	**F:**CCCCAATGCTTTTTCCATTT **R:**GGGCCTAAAATTTCTTCTTTGC	T.D (65°C - 60°C)	None
16	**F:**AAAGGCCACTCTTTCAGCAA **R:**TTCCCTGGAGCACAGAAAGT	T.D (65°C - 60°C)	None

Table 6. Summary of Variants Identified in ME2. *ME2 (NM_002396, NP_002387) was sequenced in 77 familial PCC/PGL patients. The primers used and their conditions are included in columns 2 and 3. Only one known polymorphism was identified, and is included in purple. Frequency data for the known polymorphism (rs649224) can be found in supplementary table 2.*

After analysing the electropherograms of all 77 familial PCC/PGL samples, the final analysis revealed that apart from the original nonsense mutation (p.Arg270*) identified in the exome sequencing data, and one reported common polymorphism (p.Gly450Glu), no further variants were identified.

4.3 Discussion of Familial Pheochromocytoma and Exome Sequencing

4.3a UBE2C and Pheochromocytoma

4.3a.i The Ubiquitin Pathway and Cancer

UBE2C and UBE2QL1 play key roles in the ubiquitination process; a system which is essential to the function of numerous cellular processes. Cell growth and proliferation for example, are regulated by the ubiquitin-mediated degradation of tumour suppressors, proto-oncogenes and various components of signal transduction pathways (Hershko & Ciechanover 1998). As previously mentioned, the ubiquitination process requires the functional combination of three enzymes; ubiquitin activator enzymes (E1), ubiquitin-conjugating enzymes (E2), and t ubiquitin ligase enzymes (E3).

The first step in ubiquitination is the ATP-dependent activation of a ubiquitin molecule; this is done through the creation of a thioester bond between the active cysteine residue on the E1 enzyme and the c-terminus of the ubiquitin molecule. The activated ubiquitin

is then attached to the active cysteine residue of an E2 enzyme by another thioester bond in a subsequent reaction. The E2 then acts as an escort for the ubiquitin molecule in order to bring it within close proximity to the substrate protein. The E3 proteins then act as a platform upon which the substrate and the charged E2-ubiquitin complex can come together. The E3 then initiates the E2 to ligate the ubiquitin molecule to a specific lysine residue on the target protein (Ye & Rape 2009).

In general, the nature of the ubiquitin linkages will govern the fate and function of the target protein associated. Target proteins can become mono-ubiquitinated (where one single ubiquitin is added to 1 substrate molecule), multi-monoubiquitinated (where one ubiquitin is added to many substrate molecules) or finally, polyubiquitinated (the addition of multiple ubiquitin molecules to one substrate in a chain) (Ye & Rape 2009). In many circumstances, monoubiquitination is associated with regulating DNA repair, histone function, gene expression and receptor endocytosis (Sadowski et al. 2012). Multi-monoubiquitination was previously thought to be solely involved with receptor endocytosis; however recent evidence suggests that it can also target proteins for proteasomal degradation in some instances (Dimova et al. 2012; Sadowski et al. 2012). The addition of a poly-ubiquitin chain of four or more ubiquitins linked through the Lys48 substrate is generally a rapid mechanism to efficiently target proteins for proteasomal degradation (Pickart 1998; Sadowski et al. 2012). The versatility of the

ubiquitin pathway enables it to regulate a vast array of functions, which is why disruptions to the system can often result in diseases and malignant transformations.

Both UBE2C and UBE2QL1 are classed as E2 conjugating enzymes; the class of enzymes responsible for associating the ubiquitin molecule with the target protein substrate in the presence of the E3 ligase enzyme. All E2 conjugating enzymes contain an active core domain, which includes a catalytic cysteine residue responsible for binding to the ubiquitin molecule by thioester bond (Xie et al. 2014). It is therefore conceivable to hypothesize that alterations to these genes may lead to deregulation of their associated ubiquitination pathway, and therefore also disruption of the molecular pathways in that require functional UBE2C or UBE2QL1; these disruptions may then lead to tumourigenic transformation.

4.3a.ii UBE2QL1 as a Pheochromocytoma and Paraganglioma Susceptibility Gene

Although little is known regarding the function of UBE2QL1, it has been shown to be homologous to other E2-ubiquitin enzymes, like UBE2C. In addition, it has been shown to contain a catalytic cysteine residue that binds to ubiquitin in a similar manner to other E2 enzymes (Wake 2013). UBE2QL1 was shown to act as a tumour suppressor, and it was found to be downregulated in renal cell carcinoma (Wake et al. 2013). In addition, Wake showed that interestingly, cells that stably expressed UBE2QL1 displayed a significant decrease in mTOR activation, and it was therefore suggested that the

tumourigenic activity of inactivated UBE2QL1 may occur through a mechanism involving the activation of the mTOR pathway.

It is well known that the mTOR pathway is often activated in PCC/PGL tumours, and so the suggestion that UBE2QL1 may regulate its activation seemed interesting when considering it as a potential PCC/PGL susceptibility gene. In addition, the fact that it had been associated with renal cell carcinoma was of particular relevance due to the known overlap of genes associated with PCC/PGL and renal cell carcinoma. Although no variants in UBE2QL1 had been identified in the exome sequencing data, it seemed an interesting candidate to investigate further in a larger cohort of patients.

UBE2QL1 was screened using primers provided by Dr. Wake in a cohort of 38 patients with inherited PCC/PGL. No alterations were identified from the mutation screening process; however, there are several reasons as to why this could be. Wake 2013 had also observed a lack of intragenic mutations in *UBE2QL1,* despite finding it to be largely down-regulated in renal cell carcinoma. Wake suggests that the lack of intragenic mutations may be due to the fact that disruptions to regulators or components in pathways acting upstream may result in its deregulation rather than a loss of function mutation within the coding region of the gene itself. In addition, Wake mentioned that *UBE2QL1* does not contain many coding region polymorphisms, and suggests that this

could be due to the fact that it has undergone strong negative selection during evolution, thereby indicating that it has an important physiological role that is essential to life.

Therefore, although it was an interesting gene with the potential to be related to PCC/PGL susceptibility, the results from the mutation screening in 38 patients suggests that intragenic mutations in *UBE2QL1* are not likely to be commonly associated with inherited PCC/PGL; although it is possible that studies in a larger sample size may reveal associations of lower frequency. Nevertheless, it would be interesting to know whether the expression of UBE2QL1 is down-regulated in PCC/PGL tumours as it is in renal cell carcinoma tumours. This would suggest that although not playing a direct causative role, it is possibly contributing to tumour progression through a pathway potentially involving mTOR activation. This could also indicate the presence of upstream alterations that might be reducing *UBE2QL1* expression, as suggested by Dr. Wake.

4.3a.iii UBE2C as a Pheochromocytoma and Paraganglioma Susceptibility Gene

UBE2C has been shown to be an exclusive partner of the anaphase promoting complex (APC); an E3 ubiquitin ligase that marks various cell cycle proteins for degradation via the 26s proteasome (Xie et al. 2014). *UBE2C* is essential for the destruction of mitotic cyclins and various other mitosis-related substrates involved in cell cycle progression. Interestingly, *UBE2C* is degraded by a process of autoubiquitination, which is dependent

on a destruction-box motif that is located between residues 129-132 (Xie et al. 2014). Although this alteration likely results in nonsense-mediated decay and non-expression of the protein product, it was interesting that the nonsense (p.Gln138*) was located in close proximity to the destruction box motif.

Although this alteration was predicted to disrupt protein function, there have been many cases in which genes have been found to play dual oncogenic and tumour suppressive roles in cancer, and therefore, the possibility that potentially activating mutations may be involved was also considered.

Literature has suggested that *UBE2C* has been found to be amplified in several tumour types, although this has not been studied extensively. Regardless of this, given its functional roles in the ubiquitin pathway, and close relation to a gene found to play a role in renal cell carcinoma, it was considered an interesting candidate. It was hoped that investigations into *UBE2C* would reveal one of two potential outcomes; the first being that additional, specifically located, alterations may be identified that could lead to amplification of *UBE2C*, or alternatively, the finding of further loss-of-function alterations that would suggest a potentially conditional tumour suppressive role for *UBE2C* in PCC/PGL samples.

UBE2C was screened for mutations in a cohort of 316 patients with familial PCC/PGL (in collaboration with Mercedes Robledo *et al*). Although a variant of unknown significance was identified in the cohort, (p.Ala10Ser) it was suspected to have a benign effect on protein function, as predicted by *in silico* tools, (PolyPhen and SIFT). In addition, even if this variant was found to be pathogenic, the finding of one variant from 316 (0.3%) would suggest that alterations in *UBE2C* may only be apparent in fewer than 0.5% of cases. It was therefore concluded that alterations in *UBE2C* were unlikely to be associated with the inheritance of PCC/PGL.

4.3b ERF and Pheochromocytoma

4.3b.i ERF as a Candidate Gene

ERF is a ubiquitously expressed transcriptional repressor that is thought to act downstream of the RAS/ERK pathway. In the absence of ERK activity, ERF accumulates in the nucleus in its active form, where it performs its transcriptional repressive functions; however when the RAS/ERK pathway is activated, ERK directly phosphorylates ERF, causing it to translocate to the cytoplasm to resume its inactive form (Verykokakis et al. 2007). Verykokakis *et al* 2007 showed that one of the targets of nuclear ERF transcriptional repression is c-MYC; they showed that in the absence of ERK activity, ERF represses the transcription of c-MYC within the nucleus. This finding is of particular relevance because mutations in MAX (Myc-associated factor X)

have been associated with increased PCC/PGL susceptibility (Comino-Méndez et al. 2011).

Taking into consideration that ERF has been shown to be involved in regulating c-MYC transcription, it seems possible that ERF loss of function mutations could therefore result in overexpression of c-MYC and a consequential increase in the transcription of c-MYC target genes, much in the way that is proposed to be the tumourigenic pathway caused by mutations in MAX (outlined in section 4.1b.iv). The lack of concrete evidence with regard to the ways in which MAX mutations cause PCC/PGL means that these theories remain speculative; however these possibilities rendered ERF an attractive candidate gene for PCC/PGL susceptibility.

4.3b.ii Mutation Screening of ERF

The coding regions of *ERF* were screened in a cohort of 77 familial PCC/PGL patients. Due to the large amount of DNA required for numerous reactions, the DNA samples were whole genome amplified for use in routine PCR and sequencing reactions. Performing these whole genome amplification reactions is clearly beneficial with regard to conserving quantities of valuable DNA samples; however the reactions are not without their disadvantages. A reaction which aims to replicate an exact DNA sequence multiple times will inevitably be subject to errors, as was discovered in one patient who had seemingly acquired a splicing alteration (c.373-1 G>A), 1 nucleotide upstream of

exon 4. It is statistically, highly unlikely that the whole genome amplification process would remove a variant of interest, and therefore although these occasional acquired false positive alterations were a hindrance, the whole genome amplification process enabled large numbers of reactions to be performed whilst minimising the depletion of valuable DNA stocks; thus it was still regarded to be a valuable technique which allowed maximum efficiency with regard to this study.

In general, mutational hotspots are most often found in oncogenes and not tumour suppressors, and therefore, the finding of multiple identical alterations (26 individuals with p.Glu470Val) seemed suspicious. Consistently, this alteration was also identified in a panel of ethnically matched healthy controls; and therefore, this indicated that this alteration was likely to be an unreported common polymorphism. It was therefore decided that the p.Glu470Val alteration was unlikely to be contributing to the disease in these patients. The reasons as to why these variants are currently absent from the mutational databases remains unknown; however, given the large numbers of unreported SNPs which are found to occur within Guy's hospital's in-house data of the exome sequencing data received, it is justifiable to assume that there could be many unreported SNPs that have yet to be added to the SNP databases.

In conclusion, the only variant of unknown significance identified was a synonymous variant (p.Leu81Leu); however, this variant was not in close proximity to any intron-

exon boundary, and so it was unlikely to have any functional effects to the ERF protein. In the absence of any further mutational evidence identified from *ERF* within the PCC/PGL cohort of 77 patients, it was decided that this gene was unlikely to be playing a direct role in the susceptibility of inherited PCC/PGL; however, given the roles of ERF in regulating the transcription of c-MYC, it would be interesting to know whether ERF function is disrupted in tumours overexpressing MYC.

4.3c ME2 and TCA-Cycle Mutations and Pheochromocytoma

In the 1920's Otto Warburg made the observation that tumour cells exhibited abnormally high levels of glycolysis and lactate production, even in the presence of oxygen; a phenomenon termed "The Warburg Effect" (Warburg 1956; Racker 1974). Since then, research has shown that functional mitochondria are essential to tumour progression. This has also been demonstrated in studies using ethidium bromide, a known inhibitor of mitochondrial DNA transcription, to deplete mitochondrial DNA from tumour cells. These experiments showed that the depletion of functional mitochondria in tumour cells diminished their tumourigenic phenotypes; and in particular, the cells lost their ability to grow in an anchorage-independent fashion and became more sensitive to cytotoxic drugs (Cavalli et al. 1997). Furthermore, the tumourigenic phenotype was shown to return following the transfer of fully functional mitochondria into the cells, which strongly linked the necessity of functional mitochondria for tumour progression.

The components of the tri-carboxylic acid cycle (TCA-cycle or Krebs cycle) are of particular interest as this pathway couples glycolysis occurring in the cytosol and oxidative phosphorylation occurring in the mitochondrion; in addition, many of the enzymes and proteins involved in the TCA-cycle have been shown to be associated with various tumour predispositions.

The TCA-cycle can be described as a pathway that aims to generate energy by metabolising carbohydrates, lipids and amino acids. In brief, prior to entry into the TCA-cycle, pyruvate generated from glucose during glycolysis is oxidised to form acetyl-coA. The acetyl-coA combines with oxaloacetate to form citric acid; this reaction is catalyzed by citrate synthase. The enzyme aconitase then utilises the newly formed citric acid in a reaction to form isocitrate, which in turn is converted to α-ketoglutarate by the enzyme isocitrate dehydrogenase. The α-ketoglutarate is then converted to succinyl-CoA in a reaction catalyzed by α-ketoglutarate dehydrogenase. The succinyl-CoA is converted to succinate, which is then oxidised to fumarate by succinate dehydrogenase (SDH). Fumarate hydratase hydroxylates fumarate in order to produce malate. At this point, malic enzymes, including ME2, catalyse the oxidative decarboxylation of malate back into pyruvate, which then leaves the cycle as a source of potential energy. Finally malate is converted into oxaloacetate by malic dehydrogenase to complete the cycle; this whole process is outlined in figure 21.

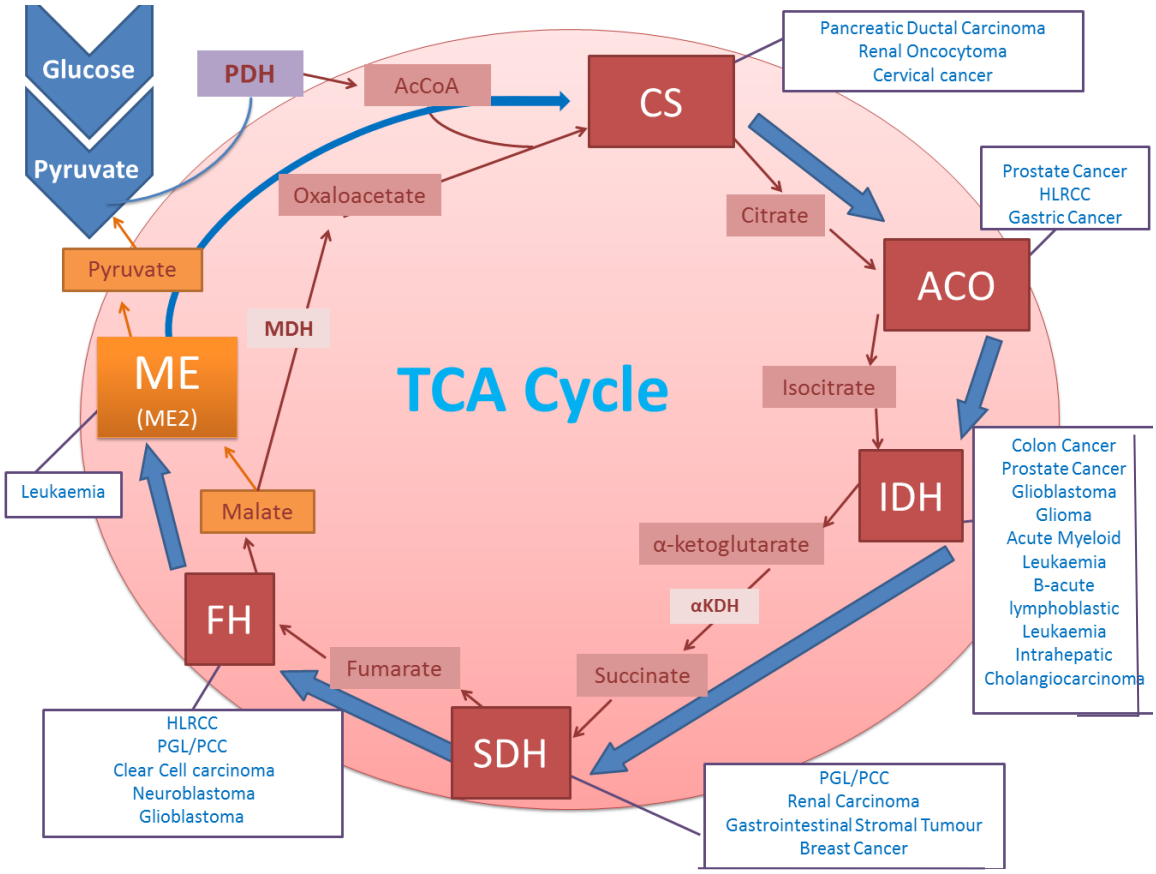


Figure 21. A Summary of the TCA Cycle *This is a schematic summary of the TCA cycle including mitochondrial enzymes known to be associated with cancer (adapted from Gaude & Frezza 2014). The boxes coloured in orange represent parts of the cycle where ME2 is involved. Abbreviations include: CS (citrate synthase), Aco (aconitase), IDH (isocitrate dehydrogenase), αKDH (α-ketoglutarate dehydrogenase), SDH (succinate dehydrogenase), FH (fumarate hydratase), ME (malic enzyme), MDH (malate dehydrogenase), PDH (pyruvate dehydrogenase), HLRCC (hereditary leiomyomatosis and renal cell cancer), PGL/PCC hereditary paraganglioma and pheochromocytoma.*

FH inactivation has been shown to act in a similar manner to SDHX inactivation, in that both succinate and fumarate accumulate and inhibit PHDs. This results in the stabilisation of HIF, consequentially generating a pseudohypoxic environment in which HIF target genes are transcribed (Gottlieb & Tomlinson 2005; Pollard et al. 2005; Selak et al. 2005). Bardella *et al* 2012 demonstrated that loss of FH function in human renal cells indeed displayed HIF1-α stabilisation and exhibited a pseudohypoxic environment; however in addition, the cells were also resistant to apoptosis, even when exposed to various apoptotic stimuli. Bardella *et al* 2012 mentions, that although upregulation of HIF has been described in FH-defective tumours, this does not explain how FH-defective cells avoid apoptosis; particularly as homozygous germline mutations in *FH* have been shown to be lethal to developing neurons (Bourgeron et al. 1994).

Their experiments showed that contrary to their original beliefs, the apoptotic resistance was occurring by a mechanism that was independent of both HIF-1α and HIF-2α. Bardella *et al* 2012 discovered that the apoptotic resistance was mediated by the activation of AMP-activated protein kinase (AMPK). This finding was surprising, particularly as AMPK has historically been reported to act as a tumour suppressor by regulating cellular energy homeostasis and suppressing cell growth and proliferation through the mTOR pathway (Hardie 2011). Bardella *et al* 2012 demonstrated that apoptotic resistance in FH-deficient renal cells was not affected by deregulating either HIF1-α or HIF2-α; however when AMPK was deactivated, the FH-deficient cells were

no longer resistant to apoptosis. These results are likely to be representative of FH-related pheochromocytomas and paragangliomas, and as mentioned previously, there is a degree of overlap of these tumours and renal cell carcinomas with regard to their susceptibility genes and methods of tumourigenesis.

To determine whether an increase in AMPK activation could be responsible for tumourigenesis, Bardella *et al* 2012, also showed that the AMPK activation in FH-deficient renal cells occurred independently of HIF stabilisation and was a direct result of fumarate accumulation; furthermore *in vivo*, they also showed AMPK was indeed activated in both renal cysts present in FH-deficient mice and also in papillary renal cell carcinoma samples from Hereditary leiomyomatosis and renal cell cancer (HLRCC) patients (Bardella et al. 2012).

Malic enzymes are tetrameric proteins that are activated by the binding of fumarate and inhibited by the binding of ATP (Chang & Tong 2003). Their main function is to catalyse the oxidative decarboxylation of malate to pyruvate while simultaneously causing the reduction of NAD(P)+ to NAD(P)H. There are three different isoforms of malic enzymes, including ME1 which resides in the cytosol, ME2 and ME3; both of which reside in the mitochondria (Zheng & Fisher 2015). The malic enzymes are evolutionarily conserved and exhibit comparable structural similarities across different species, which indicates that these enzymes play significant biological roles. Although

information about malic enzymes regarding their structural and functional aspects at a molecular level have been documented (Chang & Tong 2003), there remains much to be elucidated regarding the physiological significance and biological roles that malic enzymes play, particularly in the context of cancer. Some studies have reported observing increased expression of ME2 in certain tumours (Chang et al. 2014), which seems to contradict the theory that loss of function mutations in *ME2* may increase susceptibility to pheochromocytoma and paraganglioma. Chang *et al* 2014, for example, reported finding increased expression of ME2 in cases of cutaneous melanoma; however more importantly, and of particular relevance to inherited PCC/PGL given the effects of FH-inactivation, they found that ME2 negatively regulated AMPK activity. Chang *et al* 2014, found that AMPK activation was increased when ME2 expression was depleted, which suggests that interestingly, loss of ME2 invokes the same increase in AMPK as is found in FH-deficient tumours, and therefore, cells harbouring loss of function ME2 alterations may also gain the abilities of apoptotic resistance.

The dual roles of AMPK have been discussed in a review by Liang & Mills, 2013, and they suggest that AMPK may act as a conditional tumour suppressor or oncogene. It is possible therefore, that within tumours that acquire predispositions from mutations in TCA-cycle enzymes, that AMPK activation plays an oncogenic role in the transformation of cells, as displayed by the increase in AMPK activation observed in FH-deficient tumours. The unexpected oncogenic roles of AMPK in renal cell

carcinoma are particularly interesting when considering *ME2* as a potential causative gene, as it is possible that the process occurring in in renal tumours will reflect the activity in PCC/PGL. Therefore, as ME2 is closely related to FH within the TCA-cycle and both have been shown to increase AMPK activity when depleted suggests that germline mutations in *ME2* may act in a similar manner to germline *FH* mutations. The hypothesis is therefore, that loss of function mutations in *ME2* may lead to the disruption of the TCA-cycle and promote AMPK activation resulting in apoptotic resistance and tumourigenesis of pheochromocytomas and paragangliomas.

Despite these convincing associations, investigations into *ME2* did not uncover any further mutations other than the original nonsense (p.Arg270*) identified from the exome sequencing data; however, it should be noted that the sample size used for screening consisted of 77 patients. In comparison, the study by Castro-Vega *et al* 2014, carried out mutation screening of the *FH* gene in 598 patients, identifying 5 pathogenic alterations; a frequency of 0.8%. It is therefore not unreasonable to believe that mutations in *ME2* may still confer increased susceptibility to pheochromocytoma and paraganglioma, and the lack of further mutational evidence may be due to the limited sample size. Given the similarities between FH-depletion and ME2-depletion with regard to AMPK, it seems a likely possibility.

Little is known about the roles of malic enzymes with regard to cancer biology, and so reports of ME2 amplification in some tumours cannot eliminate the possibility that ME2 depletion may also have a conditional tumourigenic role in different tumours, especially as AMPK has been shown to play contradictory dual roles in different tumours. Furthermore, information in the TCGA database shows a high percentage of deletions and mutations in *ME2* for a variety of tumours, and only a small percentage of amplifications (figure 20), suggesting that ME2 is more often dysfunctional in tumours than overexpressed. The recent surprising oncogenic associations with the previously assumed tumour suppressor, AMPK in FH-associated renal tumours makes it a fair assumption that *ME2* may be playing a larger role in some tumours than is currently thought. Therefore, it would be of great benefit to investigate *ME2* in a cohort large enough to confidently identify mutational frequencies of less than 1% of the sample population.

Chapter Five:

SPORADIC PHEOCHROMOCYTOMA AND PARAGANGLIOMA

5.1 Introduction – Sporadic Cancer

5.1a Genetic Landscape of Sporadic Pheochromocytoma and Paraganglioma

As previously mentioned, it is thought that up to 60% of the patients who develop PCC/PGL carry either a germline or somatic mutation in a known PCC/PGL gene; in addition, many of these tumours occur sporadically without any clear familial history (Crona et al. 2013). The genetic landscape of a large proportion of many sporadic tumours without a germline mutation remains elusive. In contrast to sporadic RCC, where the majority of cases harbour somatically acquired alterations in *VHL*, sporadic PCC/PGL tumours rarely contain somatic VHL mutations. Of sporadic PCC/PGL, approximately 20-25% may harbour somatic alterations in *NF1*, 5% are found to carry somatic alterations in *RET*, and another 5% exhibit somatic *HRAS* alterations (Burnichon et al. 2012; Burnichon et al. 2011; Crona et al. 2013; Dahia 2014).

Although both inherited and somatic mutations in PCC/PGL susceptibility genes are usually found to occur in a mutually exclusive pattern, there have been some rare instances in which two somatic mutations have been found to occur in two different susceptibility genes in the same tumour (Burnichon et al. 2012). Thus Burnichon *et al* 2012 identified cases in which a single tumour exhibited somatic alterations in both *NF1* and *RET*, and another instance in which somatic *VHL* alterations were found to occur with somatic *NF1* variants. It is possible that tumour cells harbouring two mutations in

different susceptibility genes may possibly provide a growth advantage, which may exacerbate both the phenotype and prognosis. It is also possible, as previously described in section 1.3c, that these two mutations may be representative of two independent clonal expansions, that each provide separate growth advantages to the cell, and therefore may have undergone positive selection at different tumoural time points.

Studies into the genetic landscape of both hereditary and somatic PCC/PGL tumours has to date mostly focused on explaining the contributions of the highly penetrant germline susceptibility genes, and there has been relatively little research into roles of somatic alterations in PCC/PGL tumourigenesis (Luchetti et al. 2015). It is therefore clear that, although genetic research into PCC/PGL tumours has provided a wealth of knowledge and understanding with regard to tumour biology and tumourigenesis, there still remains much to be elucidated.

This part of the study aimed to shed light on the potential drivers of sporadic PCC/PGL tumours with the use of exome sequencing. It was hoped that by analysing the full exomes of multiple PCC/PGL tumours, that commonly altered genes or aberrant pathways may be highlighted that might be contributing to tumour progression. The successful clarification and interpretation of the complex genetic makeup of sporadic PCC/PGL tumours may enhance current understandings of sporadic tumour biology and pathogenesis, and most importantly could potentially be of clinical benefit to patients.

5.2 Results

5.2a Patient Samples

Tumour DNA from 13 unrelated patients with apparent sporadic pheochromocytoma and paraganglioma was sent for exome sequencing (PCC= 8, PGL=5). In addition, two tumour DNA samples from two unrelated patients with inherited paraganglioma were also sent for exome sequencing. Three of the total 15 tumour samples contained known *SDHB* mutations; the remaining 12 samples were mutation negative for any of the known PCC/PGL susceptibility genes.

As described above, some single tumours have previously been found to harbour two somatic alterations in two known PCC/PGL genes (Burnichon et al. 2012); for this reason, sequencing tumour DNA known to carry pathogenic *SDHB* mutations should not have hindered our abilities to detect further tumour-associated mutations. Furthermore, as previously described, PCC/PGL patients who carry *SDHB* mutations are often associated with more aggressive tumours, higher risks of malignancy and recurrent disease; therefore, sequencing these tumours may reveal further mutations that may be contributing to metastatic disease.

For each sample, approximately 3-5µg of tumour DNA was sent to the Biomedical Research Centre at Guy's Hospital, where they performed exome sequencing. As before,

the group in Guy's hospital identified and quality filtered all single nucleotide substitutions and small insertion deletions using a combination of the SamTools software package and their in-house software tools. They then annotated the variants with respect to genes and transcripts using the Annovar tool, and provided an excel spreadsheet containing all information.

A panel of 55 tumour DNA samples (PCC=30, HNPGL=20, PGL=5) was used for mutation screening of candidate genes selected from the exome sequencing data. Corresponding normal blood DNA was available for 34 of these samples. From the panel of 55 patients, 7 were known to carry pathogenic alterations in known PCC/PGL susceptibility genes, and a further 5 were known to carry variants of unknown pathogenicity in a known PCC/PGL susceptibility gene. The remaining apparently sporadic 43 tumour DNA samples did not carry any mutation in a known susceptibility gene.

5.2b Gene Prioritization

The exome sequencing data of 13 patients with apparently sporadic PCC/PGL and 2 patients with inherited PGL was combined into a single document in order to view all variants and generate comparisons between the samples. In total, from all 15 samples, prior to filtration, there were 394,673 variants. It was therefore necessary to apply a

series of filters to the data in order to exclude unlikely variants and retain variants of interest, which may be contributing to tumour progression.

Within the data, there were 188,337 synonymous and non-coding alterations and so these were excluded first as they are unlikely to have any effect on protein function. As was done previously, all variants with a MAF >1% were also excluded, as variants that are common within the general population are unlikely to be damaging and therefore unlikely to contribute to tumour progression. Removing all common variants reduced the number of candidates from 206,336 to 25,409; of these remaining variants, 2706 were described as novel. Following this, variants were then disregarded if they had been found over 10 times within the in-house findings in order to exclude as many false positives and unreported common polymorphisms as possible, while retaining potential variants of interest. This step reduced the number of variants to 9400 potential candidates. Due to the extremely large number of potential candidates, more stringent filtration criteria was then applied that selected genes with a MAF ≤0.1%, in order to select very rare variants within the general population. This final filter reduced the number of variants down to 3884 candidates.

Following the first stages of variant filtration, the next strategy was not to remove any alterations, but to apply a series of selection criteria in order to determine which genes are most likely to be contributing to tumour progression out of the potential 3884

possibilities. The remaining candidates were then scrutinised in order to ascertain which fulfilled the largest proportion of the selection criteria.

Selection criteria included genes known to be involved in cancer pathways (particularly kinase signalling, metabolic and hypoxia pathways), as these are well known to be aberrant in PCC/PGL tumours. In addition, genes were considered interesting if they displayed variants in more than one, but not more than 3 individuals. This was because it was thought that driver mutations in a particular gene may be present in more than one individual, particularly if similar pathways are being disrupted in these tumours; however identical alterations in a TSG in more than 3 individuals has a high probability of being due a false positive, or possibly an unreported common polymorphism. There is a possibility of finding identical alterations in an oncogene, and so identical variants could not be discarded; however, large numbers of identical variants are more likely to be false positives or common polymorphisms of no clinical relevance. Therefore, identical variants in a maximum of 3 individuals was considered to be a generous cut off point to exclude false positives, while retaining potential tumour-associated oncogenic variants.

Genes that were found to be highly mutated in the The Cancer Genome Atlas (TCGA) database were considered interesting, as this indicates that they may potentially have a role in tumour progression. Finally, the type of variant was also considered with

particular attention on variants that were likely to have a clear loss-of-function consequence (frameshift, nonsense, and splicing alterations that were no more than 3bp from the intron/exon boundary). Finally, when genes passed through these criteria, the gene function was considered in order to assess the likelihood of it having a potential role in tumour development or progression.

5.2c HIF Pathway and Sporadic PCC/PGL

As explained in section 4.1a, pseudohypoxia arises when HIF pathways are activated in the absence of hypoxia. The response to hypoxia is regulated by Hypoxia-inducible factors (HIFs); heterodimeric transcription factors that become activated in response to a hypoxic or pseudohypoxic environment, and they function to regulate the cell's adaptive responses by initiating the transcription of numerous target genes.

HIFs are made up of an α-subunit and a β-subunit; both of which belong to a family of transcription factors known as the basic-helix-loop-helix (bHLH) PAS proteins (Wang et al. 1995; Ratcliffe 2013). The β-subunit is a stable, constitutively expressed protein that often forms heterodimers with other members of the bHLH family in order to perform various cellular functions (Jochmanová et al. 2013). The α-subunit consists of three subunits, HIF-1α, HIF-2α and HIF-3α, and is solely involved in the hypoxic response. HIF-α is largely regulated by prolyl hydroxylases (PHDs) that hydroxylate the subunits at specific proline residues under normoxic conditions. This process of hydroxylation

then enables pVHL, the recognition component of a tumour suppressor complex with E3-ubiquitin ligase activity, to recognise the substrates and initiate their subsequent ubiquitination through proteasomal degradation (Maxwell et al. 1999; Kaelin & Ratcliffe 2008). Following the discovery of the association of pVHL and HIF, it was soon suggested that the angiogenic tumourigenesis of VHL disease-associated tumours may occur through a mechanism involving the stabilisation of the HIF-α subunit and an accumulation of HIF-1 as a direct consequence of VHL inactivation, thereby resulting in the subsequent transcription of a multitude of target genes, including those involved in angiogenesis.

In keeping with these predictions, both *VHL* and *SDHX* mutations have been shown to result in an accumulation of HIF and increased HIF activity (Maxwell et al. 1999; Selak et al. 2005). Given the roles of both *VHL* and the *SDHX* genes in PCC/PGL tumourigenesis, it seemed reasonable to hypothesise that specific activating mutations directly affecting the HIF complex might reciprocate the same consequences as *VHL* and *SDHX* mutations, and therefore may also contribute to PCC/PGL tumour development and progression.

Multiple studies have outlined the role of dysregulation of HIF2-α in tumour progression; but up until recently, there had been no previous reports of mutations in *HIF2A* directly having an association with a particular neoplasm. The first publication

that reported this was released during this PhD investigation; it described the identification of gain-of-function somatic alterations in *HIF2A* in patients with paraganglioma associated with congenital polycythaemia (Zhuang et al. 2012). The wealth of functional relevance of HIF-2α in PCC/PGL associated pathways, in addition to the release of recent publications outlining its association with these tumours initiated the decision to carry out further investigations into the involvement of *HIF2A* mutation in PCC/PGL tumours in the hope to gain a deeper understanding of its role in tumourigenesis.

HIF-2α is hydroxylated by PHDs at 2 specific proline residues; one is located in exon 9, and the other in exon 12, and therefore it can be inferred that variants occurring within these regions are likely to alter HIF-2α function. Interestingly, the exome sequencing data revealed two novel missense alterations in *HIF2A*; one in exon 9 in an individual with paraganglioma (p.Ser372Asn) and the second in exon 12 in an individual with pheochromocytoma (p.Pro531Thr). Interestingly, the p.Ser372Asn alteration identified in the exome sequencing data was found to be in an individual that also carried a pathogenic *SDHB* mutation (p.Trp200Cys).

A publication by Zhuang *et al* 2012 released at the time of these findings also identified the same p.Pro531Thr variant, which provided a further indication that this particular variant was playing a pathogenic role in the associated individuals. It was therefore

decided that exons 9 and 12 should be screened in the cohort of 55 PCC/PGL tumour DNA samples in order to determine whether any further novel alterations were present that could potentially be driving tumour progression.

5.2d Mutational Screening of HIF2A in Sporadic PCC/PGL

Two sets of primers were designed to flank exon 9 and exon 12 respectively. These regions were PCR amplified primarily in the individuals implicated in the exome sequencing data to confirm the presence of their variant by Sanger sequencing. The results revealed that both p.Pro531Thr and p.Ser372Asn were true variants and present in the non-amplified tumour DNA of the associated individuals (displayed in figure 22).

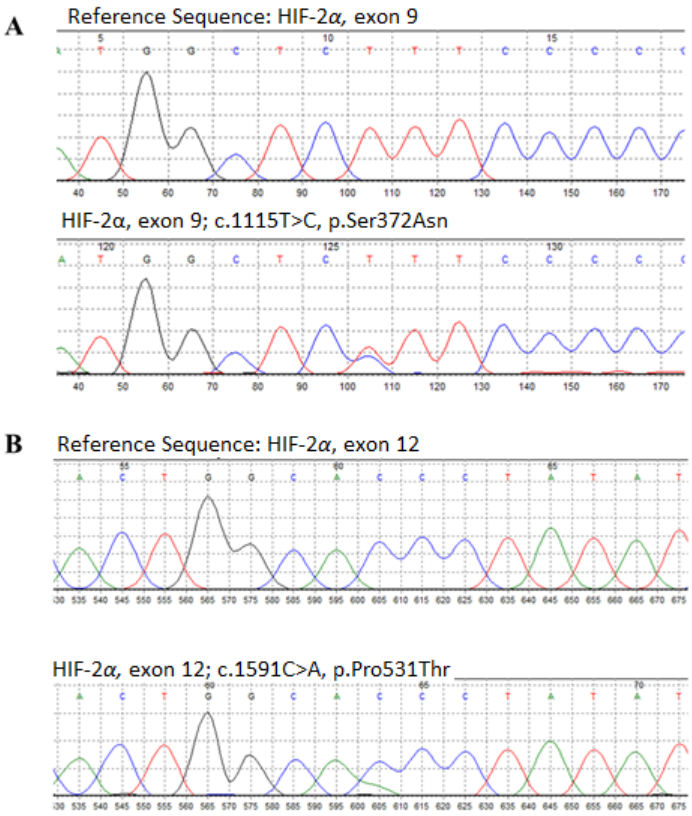


Figure 22. ***Confirmation of Variants Identified in HIF2A:*** *A: Shows a reference sequence for a specific region in HIF-2α, exon 9 and the p.Ser372Asn alteration identified. B: Shows a reference sequence for a specific region in HIF-2α, exon 12 and the p.Pro531Thr alteration identified.*

Mutational screening was then carried out in a cohort of 55 whole genome amplified PCC/PGL tumour DNA samples. Interestingly, in exon 9, one additional missense variant p.Pro407Arg was identified in the same individual that carried the p.Pro531Thr alteration (figure 22). This second identified variant was re-tested in the original stock DNA of this sample, and found to be a true alteration.

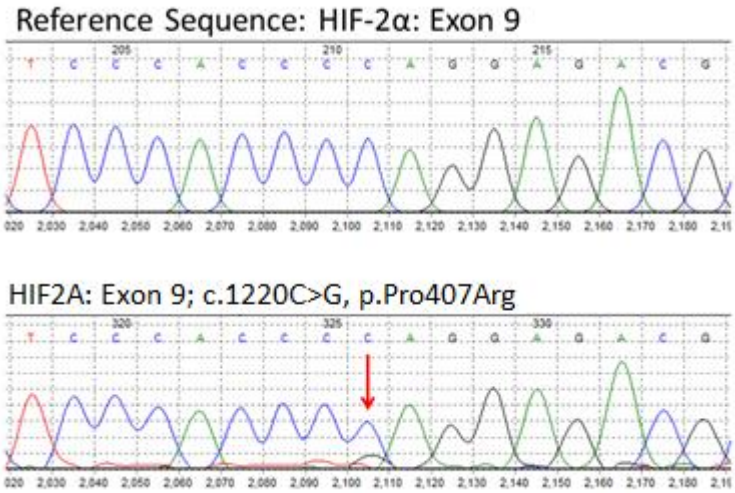


Figure 23. Confirmation of p.Pro407Arg Identified in HIF2A Exon 9 *This image shows a reference sequence for HIF2A exon 9, and the p.Pro407Arg alteration identified. The peak of the altered allele is fairly low level, which could be indicative of mosaicism.*

Mutation screening of exon 12 revealed one further variant (p.Phe583Leu) in one individual with sporadic paraganglioma. This variant was confirmed in the original stock DNA of that tumour sample. In addition, corresponding normal blood DNA was available for this sample and so this was also tested for the presence of the variant. The results revealed that the variant was present in the blood DNA sample, confirming that the variant was not an acquired somatic alteration. However, it does appear that the variant allele is enriched in the tumour DNA, as the peak for the variant allele is higher in the electropherogram of the tumour DNA in comparison to the peak found in that of the blood (figure 24).

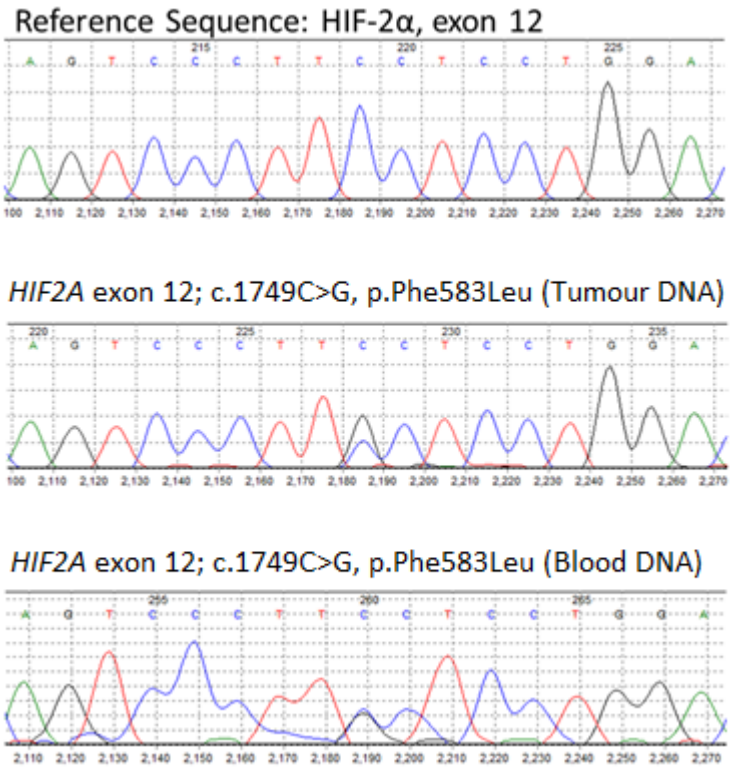


Figure 24. Electropherograms of HIF-2α Exon 12. *This image shows a reference sequence for HIF-2α exon 12, the p.Phe583Leu alteration identified in the tumour DNA and the same alteration found in the corresponding blood DNA.*

Therefore, in total, from the panel of 55 tumour DNA samples (48 assumed sporadic), four alterations were identified in *HIF2A* in three individuals. Two alterations were identified from the exome sequencing data, and two further variants were identified from the mutation screening of exons 9 and 12. A summary of these variants is outlined in Table 7.

Sample	Alteration	Exon	Identification Method	Somatic Status	PolyPhen	Previously Reported
1T	p.Pro531Thr	12	Exome	N/A	Probably Damaging	Yes
1T	p.Pro407Arg	9	Sanger	N/A	Predicted Benign	No
2T	p.Ser372Asn	9	Exome	N/A	Predicted Benign	No
3T	p.Phe583Leu	12	Sanger	Germline	Predicted Benign	No

Table 7. ***Summary of Alterations Identified in HIF2A From a Cohort of 55 Apparently Sporadic PGL/PCC Samples.*** *Four alterations were identified in three individuals. The p.Pro531Thr alteration, which had previously been reported in the publication by Zhuang et al 2012, was predicted to be probably damaging by PolyPhen. The remaining three novel variants were all predicted to be benign. It was not possible to determine the somatic status of the first three variants, as there was no constitutional blood DNA available for these samples; therefore they are marked as N/A.*

5.2e HRAS

As previously mentioned in section 4.1b, there are two major molecular pathways that are associated with the formation of PCC/PGL tumours, enabling them to be separated into two distinct groups based on the genetic mutations present. Cluster 1 involves

aberration of the hypoxic pathway while cluster 2 includes tumours characterised by abnormal activation of the PI3K-AKT-mTOR and RAS/ERF/ERK kinase signalling pathways. Cluster 2 tumours are most often associated with mutations in *RET, NF1, MAX* and *TMEM127*; all of which cause activation of the kinase signalling pathways mentioned.

The mammalian target of rapamycin (mTOR) is an evolutionarily conserved serine/threonine kinase that becomes activated in response to various mitogenic stimuli. This then initiates the activation of numerous cellular metabolic pathways that promote cell cycle progression and drive cell proliferation (Shimobayashi & Hall 2014). The mechanism of tumourigenesis in cluster 2 therefore often involves mutations that cause aberrant, constitutive activation of the mTOR pathway, ultimately leading to increased cell growth and tumour development.

The RAS family of proteins consists of HRAS, KRAS and NRAS; these proteins are GTPases that act as molecular switches. They largely function by coupling cell surface receptors to intracellular signalling effector pathways in order to promote cell growth, proliferation and survival (Pylayeva-Gupta et al. 2011). RAS proteins are auto-regulatory as they have the intrinsic ability to initiate GTP hydrolysis; this enables them to cycle between an active GTP-bound state and an inactive GDP-bound state.

RAS proteins are among the most commonly mutated genes in cancer, and are known to contain specific oncogenic mutational hotspot regions. These oncogenic hotspots include the residues Glycine-12 (Gly12), Glycine-13 (Gly13) and Glutamine-61 (Glu61) (Gripp & Lin 2012). Glu61 is a catalytic residue that is essential for GTP-hydrolysis and a substitution at this location for any amino acid other than glutamine will cause a conformational change that prevents GTP-hydrolysis from occurring. This prevents RAS from cycling into the GDP-bound state, and therefore results in a constitutively active RAS pathway that continuously initiates a pro-tumourigenic downstream signalling cascade (Scheffzek et al. 1997; Pylayeva-Gupta et al. 2011). Substitutions occurring at residues Gly12 and Gly13 prevent the association of GTPase Activating Proteins (GAPs). GAPs are proteins which initiate the activation of GTP hydrolysis, and for this reason, loss of GAP association will also prevent RAS from cycling into the inactive GDP-bound state, which also results in a constitutively active RAS pathway (Pylayeva-Gupta et al. 2011).

During the course of this study, a paper was released which reported finding mutations at these key residues in *HRAS* associated with sporadic cases of PCC/PGL (Crona et al. 2013). Interestingly, upon investigation, one individual in the exome sequencing data from the sporadic PCC/PGL tumour DNA samples in our study was found to carry one of the same missense mutations that was outlined in the paper (p.Gly13Arg). This

finding then raised the question as to whether additional variants in the RAS family might also be driving tumour progression in sporadic PCC/PGL tumours.

In order to investigate this, primers were designed to flank exonic hotspot regions of *HRAS, KRAS* and *NRAS*. The primers were used to PCR amplify these regions in a cohort of 55 tumour DNA samples from patients with apparently sporadic PCC/PGL. The samples then underwent Sanger sequencing in order to determine if any variants of interest were present. In addition to the original p.Gly13Arg variant present in the exome sequencing data, three samples were identified that carried the substitution of p.Glu61Arg; this in total amounted to 4 individuals with activating HRAS mutations; one with p.Gly13Arg and 3 with the p.Glu61Arg substitution. The somatic/germline status of these variants could not be confirmed as no constitutional blood DNA was available for these patients (table 8).No variants of interest were found in *KRAS* or *NRAS* in the cohort of sporadic PCC/PGL tumour DNA samples.

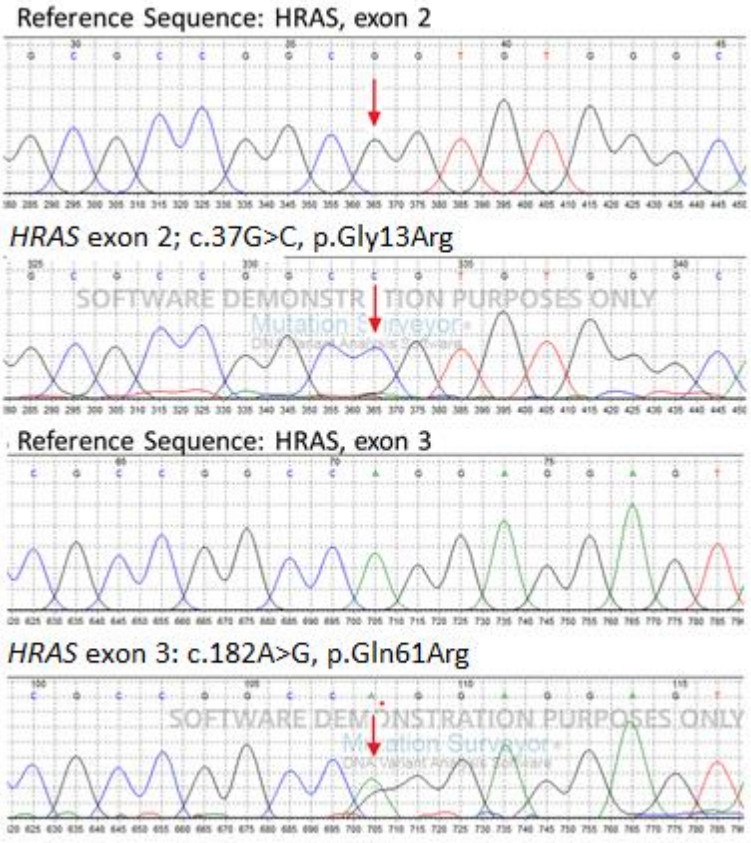


Figure 25. Electropherograms Outlining Variants Identified by Sanger Sequencing in HRAS. *The images show a region of reference sequence for HRAS, exon 2 and the corresponding region in an affected individual harbouring a p.Gly13Arg homozygous alteration. The subsequent images show a region of reference sequence for HRAS, exon 3, and the corresponding region in an affected individual harbouring a heterozygous p.Glu61Arg alteration.*

Following these findings, an opportunity arose in collaboration with Andrea Luchetti of the Department of Medical Genetics at the University of Cambridge, to investigate these samples using the IonAmpliSeq cancer hotspot panel v2 (Life Technologies, UK). This next generation sequencing panel targets the mutational hotspot regions of 50 known

cancer associated genes including *HRAS, KRAS* and *NRAS*. This method enables these regions to be sequenced in parallel with higher coverage than would be achieved with whole exome sequencing. Furthermore, this method is also more sensitive to the presence of mosaic variants than Sanger sequencing. The collaborators in Cambridge used the same cohort of 55 tumour DNA samples, and also included a further 30 tumour DNA samples with apparently sporadic PCC/PGL. Of the total 85 samples, 21 were known to carry germline pathogenic alterations in a known susceptibility gene.

In addition to the alterations identified through Sanger sequencing, IonAmpliSeq also identified two further p.Gln61Arg alterations in two more individuals. These two alterations had not been apparent in the original Sanger sequencing data; possibly due to the fact that these were mosaic alterations. This raised the total number of variants identified in *HRAS* to 6 out of a total 85 tumour DNA samples from patients with sporadic PCC/PGL; one with p.Gly13Arg, and 5 with p.Gln61Arg. This amounted to an overall frequency of 7.1%.

In addition to the variants in *HRAS*, the IonAmpliSeq data also revealed one activating mutation in *BRAF*, (p.Val60Glu), and two tumours harbouring missense alterations in *TP53* (p.Arg337His) (Luchetti et al. 2015). No *HRAS* mutations were identified in any of the 21 patients that carried a pathogenic germline mutation (Luchetti et al. 2015).

Amino Acid Alteration	Nucleotide Alteration	Gene	Somatic Status	Identified by
p.Gln61Arg	c.182A>G	***HRAS***	N/A	Sanger Seq.
p.Gln61Arg	c.182A>G	***HRAS***	N/A	Sanger Seq.
p.Gln61Arg	c.182A>G	***HRAS***	N/A	Sanger Seq.
p.Gly13Arg	c.37G>C	***HRAS***	N/A	Exome Seq.
p.Gln61Arg	c.182A>G	***HRAS***	Somatic	IonAmpliSeq
p.Gln61Arg	c.182A>G	***HRAS***	Somatic	IonAmpliSeq
p.Val600Glu	c.1799T>A	*BRAF*	Somatic	IonAmpliSeq
p.Arg337His	c.1010G>A	*TP53*	Somatic	IonAmpliSeq
p.Arg337His	c.1010G>A	*TP53*	Somatic	IonAmpliSeq

Table 8. A summary of variants identified in HRAS and other cancer genes using a combination of Sanger sequencing and the IonAmpliSeq cancer hotspot panel. *This data was published in collaboration with Luchetti et al 2015, and some information has been taken from this publication. The somatic status of some of the individuals remains unknown as the corresponding blood DNA was not available for these samples, and so they are marked as N/A.*

5.2f KEAP1 and CUL3

Following investigations into *HIF2A* and *HRAS*, the exome sequencing data for the sporadic tumour DNA samples was reviewed in order to determine whether any further

candidate genes of interest were present. Following rigorous filtration of the sporadic PCC/PGL exome sequencing data, the remaining variants were scrutinized in order to prioritise and identify likely PCC/PGL candidate genes. Genes were assessed based on their predicted functional impact on the encoded protein, with the criteria that nonsense and frameshift alterations are more likely to have a deleterious effect on protein function than missesnse alterations. In addition, genes were also analysed based on the functional relevance of their encoded proteins, in order to assess the likelihood of their potential involvement in tumourigenesis.

Following this process, one gene was noticed to be particularly interesting; Kelch-like ECH-associated protein (*KEAP1*). KEAP1 is best known for its associations with CUL3 and Nrf2 as part of a crucial adaptive response pathway to cellular stress; a pathway essential for the prevention of human diseases including cancer (Zhang 2010). It has been suggested that KEAP1 frequently exists in a complex with CUL3, and that in the absence of cellular stress, this complex associates with Nrf2 and promotes its degradation via the ubiquitin pathway (Cullinan et al. 2004; Kansanen et al. 2013)

KEAP1 contains many cysteine residues, several of which are essential to its function; namely residues Cys151, Cys273 and Cys288, as these are involved in sensing cellular stress. In the presence of cellular stress, excess concentrations of ROS for example, these cysteine residues on KEAP1 become modified, causing a conformational change

in the protein structure of KEAP1 and consequential dissociation from Nrf2. This reverses the proteasomal degradation of Nrf2, enabling it to translocate to the nucleus and initiate the activation of Nrf2 downstream target genes, which then enable the cell to adapt to the stresses that are apparent (Zhang 2010; Kansanen et al. 2013; Cullinan et al. 2004).

Genomic alterations in *KEAP1* which affect of any of the key cysteine residues Cys151, Cys273 or Cys288 have been reported to disrupt protein function by altering the conformation of KEAP1 in a similar manner to the presence of cellular stress. This results in the inability of the KEAP1-CUL3 complex to associate with Nrf2, and therefore Nrf2 is not degraded via the ubiquitin pathway but remains constitutively active (Kansanen et al. 2013). Cullinan *et al* 2004 showed that the individual role of KEAP1 is to act as an adaptor or bridge which enables the association of CUL3 with Nrf2. CUL3 then acts as an E3 ubiquitin ligase protein and initiates the ubiquitin-mediated degradation of Nrf2 (Cullinan et al. 2004). Cullinan *et al* 2004 showed that targeted inhibition of either CUL3 or KEAP1 *in vitro* lead to an accumulation of Nrf2 and an increase in the transcription of Nrf2 downstream signalling targets; this suggests that both proteins must be functional for the successful ubiquitination of Nrf2.

In addition to its roles in the cell's adaptive responses to stress, Nrf2 has been shown to have many pro-tumourigenic functions including inhibition of apoptosis, and

involvements in cellular proliferation (Martinez et al. 2014). Consistently with this, hyperactivation of Nrf2 and increased expression of Nrf2 downstream target genes has been described in numerous different cancers and is thought to occur via several different mechanisms. One example is by the presence of somatic mutations to KEAP1 that prevent Nrf2 proteasomal degradation (Hast et al. 2014). Another mechanism, of particular interest to this study, is the activation of Nrf2 caused by an accumulation of fumarate; a process that has been reported to be part of the tumourigenic pathway in FH-deficient renal tumours (Sporn & Liby 2012; Adam et al. 2011). In these cases, FH inactivation leads to an accumulation of fumarate which acts as a form of cellular stress, and therefore causes post-translational modifications to the key cysteine residues of KEAP1 by succination (Adam et al. 2011). This alters the conformational structure of KEAP1 thereby preventing the KEAP1-CUL3 complex from associating with Nrf2, and so also prevents its proteasomal degradation, resulting in constitutive activation (Kansanen et al. 2013; Cullinan et al. 2004).

The recent discovery that FH inactivation has also been shown to be associated with PCC/PGL tumours suggested a hypothesis that alterations in *KEAP1*, or *CUL3* for that matter, could potentially be involved in the development of PCC/PGL tumours. One of the key reasons for choosing *KEAP1* as a candidate gene, aside from its reported tumourigenic role in FH-deficient renal tumours was the fact a nonsense alteration was detected in *KEAP1* nonsense in the exome sequencing data for an individual with

sporadic HNPGL. Interestingly, the nonsense mutation was found to affect one of the key stress-sensing cysteine residues, (p.Cys273*). As this alteration was likely to cause nonsense-mediated decay, it was considered likelt that this variant would disrupt KEAP1 function. The hypothesis that mutated *KEAP1* was a potential driver of tumourigenesis in PCC/PGL was therefore based on a combination of the suggested involvement of KEAP1 in FH-deficient renal tumours, the fact that *FH* had recently been implicated in the development of PCC/PGL tumours, and the presence of a nonsense variant; a variant highly likely to disrupt protein function. In addition, given the close association of KEAP1 and CUL3, and the fact that Cullinan *et al* 2004 had previously shown that disruption to either protein could lead to hyperactivation of Nrf2, *CUL3* was also selected as a candidate gene that may be playing a tumourigenic role in PCC/PGL tumours.

In order to carry out mutation screening, primers were designed to flank the coding exons of *KEAP1* and *CUL3.* The primers were used to PCR amplify regions of interest in a panel of 55 tumour DNA samples from patients with PCC/PGL (30=PCC, 20=HNPGL, 5=PGL). As described previously, 48 individuals in this panel were suspected to have sporadic cases of PCC/PGL, while 7 patients were known to carry pathogenic alterations in a known PCC/PGL susceptibility gene. In addition, these regions were also amplified in the constitutional blood DNA samples from 83 patients with suspected familial PCC/PGL.

KEAP1 Exon	Primers	Conditions	Variants		Frequency
			Familial	Sporadic	
2	**F:**TATCTTGCAAAACGAGGCCC **R:**AAGGGGAGACAGTGATGAGC	Std58°C	None	None	-
3	**F:**GTCAGCGGCAGTGATAAGTT **R:**CGGATCTCAGTGTCTTGGGA	Std58°C	None	N/A	-
4	**F:**CACGAAGGTCAGCTATAATGGC **R:**GGGTGTTCCTGGGTGCTC	Std60°C	1.p.Leu471Leu	2. p.Leu471Leu 3. p.Asp479Asp	1. 12/83 2. 14/55 3. 3/55
5	**F:**GTTCCCAAAGCCAGACCC **R:**GCAAAAGCAAAAGCAGTCCA	Std60°C	None	1. p.Tyr537Tyr 3. p.Ile519Val	1. 10/55 3. 1/55
6	**F:**ATGTGGTGTGACAGGTGGTG **R:**TGCTGTCTTTTCTTTTAGTCCCG	Std60°C	1. p.Thr590Thr 2. p.Gly605Gly	None	1. 1/83 2. 1/83

Table 9. A Table Outlining The Variants Identified in KEAP1 (NM_203500, NP_987096) in the Cohort of 55 tumour DNA and 83 Constitutional Blood DNA Samples from Patients with Sporadic or Inherited PCC/PGL. *Variants in purple are known reported polymorphisms, variants outlined in orange represent unreported synonymous alterations and variants in green represent unreported variants of unknown clinical significance. More information is provided for the reported polymorphisms in supplementary table 2.*

From the combined total of 143 blood and tumour samples analysed, several common single nucleotide polymorphisms, and some synonymous alterations were identified. None of the unreported synonymous variants identified (p.Thr590Thr, p.Gly605Gly) were within close proximity to any splice sites, and it was therefore predicted that they would be unlikely to cause any detrimental effects to KEAP1 function. One heterozygous variant of unknown clinical significance was identified in a sporadic paraganglioma sample (c.1555A>G; p.Ile519Val) (figure 26). As no constitutional blood DNA was available for this patient, it was not possible to determine the somatic/germline status of the variant. Furthermore, the alteration was predicted to have a benign effect on protein function by the *in silico* mutation prediction software tools; SIFT and PolyPhen (figure 27).

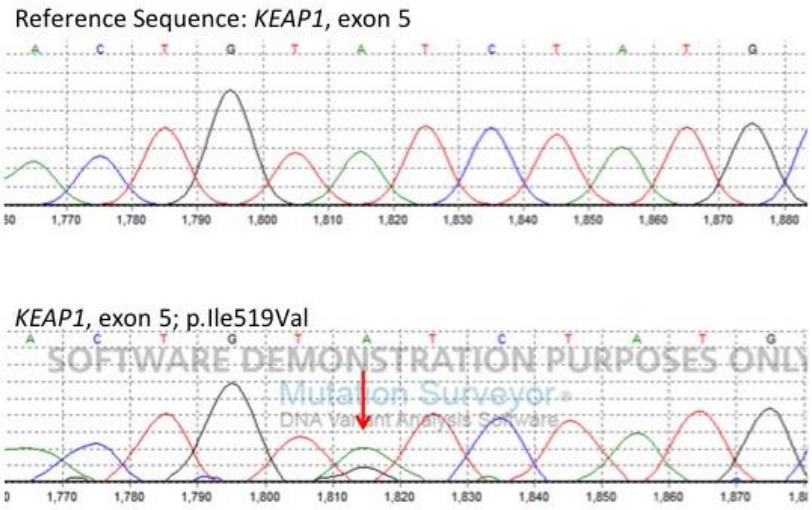


Figure 26. Electropherograms Outlining a Variant of Unknown Significance Identified in KEAP1. *The images show a region of reference sequence for KEAP1, exon 5 and the corresponding region in an affected individual harbouring the p.Ile519Val alteration.*

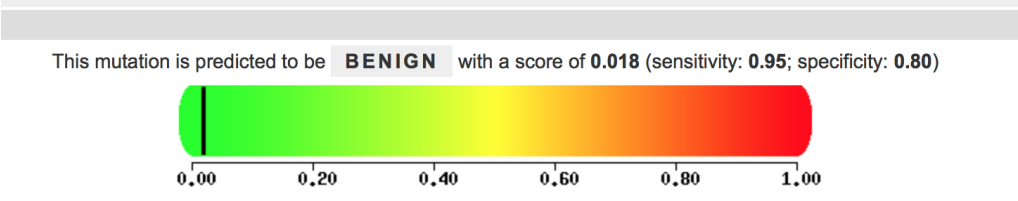


Figure 27. PolyPhen Mutation Prediction Software. *The alteration of p.Ile519Val in KEAP1 is predicted to have a benign effect on protein function. This prediction is based on a combination of factors including the degree of conservation of this amino acid across different species, and the location of the amino acid in relation to the important functional domains of the protein.*

As no further variants were found that were predicted to have any altering effect on protein function, *KEAP1* was excluded from further analysis.

All coding exons of *CUL3* were analysed in the same panel of 55 tumour DNA and 83 constitutional blood DNA samples from a combination of patients with suspected sporadic and inherited forms of PCC/PGL. The results are outlined in Table 10.

CUL3 Exon	**Primers**	**Conditions**	**Variants**		**Frequency**
			Familials	Sporadics	
1	**F:**CTGCCCGCCTTAAATGTGAC **R:**AGCCCCTTCATCACCCTAAAA	Std57°C with GC-Rich	None	None	-
2	**F:**CAGGTAAGATTGGAGGGAGGT **R:**CCGGTCAAAGTGCAATATGGT	Std60°C	None	None	-
3	**F:**TCACTGGAGGGTTGGGAAAT **R:**TGACCTTAAATGATATGCCCACC	Std53°C	None *(Sequenced by Atif Alsaedi)*	None *(Sequenced by Atif Alsaedi)*	-
4	**F:**GAGAGGTGCTGGAAAGTTGC R: GGATTTTCCCAATGTGCTCAAC	T.D 65°C - 55°C	1.p.Arg173Arg	None	1. 3/83
5	F: ATTGGAGATTGGTCACTTCAGT	Std60°C	None	None	-

	R: GTGAGGTAGTGTGCTG ATGG				
6	F: ACACTGTAGAAACCCA GCTAAA R: AGGACTAAAGAGAGC CCACA	Std60°C	1. c.685+37 C>T (DIV) 2.c.685+39C>G (DIV)	3. c.685+37 C>T (DIV) 4.c.685+39C >G (DIV)	1. 24/83 2. 31/83 3. 13/55 4. 18/55
7	F: GCGTGAACAGAGATTC GTGT R: ACAGCATGGTAAAAGT GGCC	Std60°C	None	None	-
8	F: AGAAACAGTGAGGAG GTTCTTT R: ACGTTTGGGCATGTCA AGTA	Std60°C	None	None	-
9	F: TGAACCGTAGTACTGC TTTTGT R: AGTGTAGTCTGTGTTC TTCTCCA	Std60°C	1.c.1179+48 T>C	2. c.1179+48 T>C	1. 70/88 2.36/55
10	F: TGATAACCACAGGCAC AGAG R: CAACTGTCATCTGCTA AGTGGA	T.D 65°C - 55°C	None	1.p.Ile479Ile 2. p.Thr483Thr	1. 1/55 2. 1/55
11	**F:**AAACAGAACAGAAA GCAAATCCT **R:**TGCCTGCTTTACTTT	-	N/A	N/A	-

	ACTCATCA				
12	**F:**GTGTGGGGCCTTTGTCATTT **R:**TGTGAGACAGGCAGAGAAACA	Std.60°C	1. p.Val501Ile	2. p.Val501Ile	1. 17/83 2.16/55
13	**F:**CCAGGAAATGATTGTGTCTTTGT **R:**TGCTCTGTATTTTCGGGCAA	-	N/A	N/A	-
14	**F:**TGCTTGGTCAGTTTACTTGCT **R:**CATGCAGCACCAGAAAAGGA	T.D 65°C - 55°C	1.p.Gln598Gln	2.p.Gln598Gln	1. 11/34 2. 20/55
15	**F:**ACTCTTGTCTCCTCATAAAGCCT **R:**CATTGGTTCCGAGAGTGTGC	Std 58°C	None	None	-
16	**F:**TCTGATTTGAGGAAGTGTGTTGT **R:**TGTTGGAAACTCTCAAAGGGA	Std 60°C	None	None	-

Table 10. A Table Outlining the Variants Identified in CUL3 (NM:003590, NP:003581) from a Cohort of 55 Tumour DNA and 83 Constitutional Blood DNA samples from Patients with Sporadic or Inherited PCC/PGL. *Previously reported known polymorphisms are outlined in purple, unreported synonymous variants are outlined in orange. No pathogenic variants or variants of unknown pathogenicity were identified during this screening. Some exons were not able to be completed, and therefore are marked with N/A.*

The sequencing data revealed several common polymorphisms and several deep intronic variants. However, it should be noted that, due to the time restrictions of this project, it was also not possible to sequence every exon to completion in every sample. Help was received from Mr. Atif Alasaedi to sequence *CUL3* exon 3. However, exons 11 and 13 were unable to be sequenced. It is expected that this is due to inefficient primer sets. A minimum of 20 patients were sequenced for all other exons in each cohort of familial and sporadic PCC/PGL samples (supplementary table 3). The incompletion is largely due to the fact that it is difficult to obtain successful, clean sequencing for every sample in the first round of sequencing, and it is inevitable that some samples will fail (due to factors such as incorporation of dye blobs for example, or through occasional, unavoidable human errors).

5.2g Cullin-2 and VHL

As previously mentioned, cluster 1 PCC/PGL tumours are well known to exhibit pseudohypoxic environments, often showing increased expression of hypoxia-associated target genes. This can be achieved through a variety of mechanisms including genetic inactivation of *VHL* through the occurrence of a deleterious mutation, or, as recently discovered, specific activating mutations occurring in the hotspot regions of *HIF2A* (Comino-Méndez et al. 2013). An intriguing aspect to pseudohypoxic PCC/PGL tumours is that *VHL* mutations are only identified in less than 10% of tumours, which cannot account for the high frequency with which pseudohypoxic tumour environments

are displayed (Rowbotham et al. 2014). Therefore, it is hypothesized that additional mechanisms that give rise to this phenotype may be present that have not yet been elucidated. There is currently little known about the contributory role of other components of the hypoxia pathway in PCC/PGL tumour development and progression. Therefore, given the high rate of pseudohypoxic environments in PCC/PGL tumours in combination with the disproportionately low number of inactivating *VHL* mutations or activating *HIF2A* mutations, it seemed logical to suggest that disruptions to other parts of the hypoxia pathway may be contributing to this phenotype.

The pVHL protein has been shown to exist in a complex along with elongation proteins (elongin B / elongin C), CUL2 (a member of the cullin family) and ring finger protein RBX1; which together form the VCB-CR complex (Gossage et al. 2014). This complex then acts an E3 ubiquitin ligase that promotes the targeted ubiquitination and proteasomal degradation of HIF-1α and HIF-2α under normoxic conditions (Gossage et al. 2014). Although no *CUL2* knock-out mouse models have been examined to date, previous studies have shown that targeted inhibition of CUL2 in epithelial cells leads to the stabilisation and accumulation of HIF, much in the same way as VHL (Curtis et al. 2014; Chen et al. 2015). This finding gives weight to the hypothesis that inactivating genetic mutations in *CUL2* might potentially provide an explanation for a proportion of the unexplained pseudohypoxic environments observed in many PCC/PGL tumours lacking VHL or HIF-2α alterations.

The exome sequencing data of our study contained one missense alteration in *CUL2* (c.1051C>G, p.Leu351Val) from the cohort of 12 tumour samples; however this alteration was predicted to have a benign effect on protein function by *in silico* prediction software tools, SIFT and PolyPhen. Regardless of this prediction, it seemed intriguing to investigate the potentially progressive role of *CUL2* alterations in PCC/PGL patients; particularly given the recent findings of *HIF2A* alterations in PCC/PGL tumours. Therefore, primers were designed to flank all coding exons of *CUL2*, for use in direct sequencing as part of the mutational screening process in a panel of 55 tumour and 83 constitutional blood DNA samples from familial and sporadic PCC/PGL patients.

CUL2 Exon	Primers	Conditions	Variants		Frequency
			Familial	Sporadic	
2	**F:**GACTCCGTCTCAAAGCAAACA **R:**TGCCATAAAAGAGGTGATGAACT	TD 65°C - 55°C	None	None	-
3	**F:**CCCTCTGGATTTTAACCTCTCTG **R:**GCAACAGAGCAAGACCCTG	Std53°C	None	None	-
4	**F:**CGCTCGGCTCAGTTTATTCT **R:**TAGACCTTTCCCCACAGCAG	Std60°C	1. p.Lys109Glu 2. Glu106Glu	None	1. 1/83 (benign) 2. 1/83
5+6	**F:**TGCATTTCTTCCTGGTGATTAGA **R:**GCTAGCCAAATCAGTGTCCT	Std60°C	None	None	-
7	**F:**TGTTGATAAGGCTAGTAGAGTGG **R:**TGGCTCTGCAATTGTACACT	Std60°C	1. c.613-6C>T (32/88)	2.c.613-6C>T (21/55)	1. 19/88 2. 8/55
8+9	**F:**TGATGGGCTCGTATCTACTTTG **R:**AACCAGGCCAGATACAAAATAGA	Std54°C	None	None	-
10	**F:**CCTTTGCCATGTTTAATTCAGCC **R:**TCAAGGGTGACTAGAGGATGA	TD 65°C -60°C	None	None	-

11	**F:**CTTGGGTGGCTGGGAGGT **R:**GACATTTCTTTCGTTGGCATGGT	TD 65°C - 60°C	None	None	-
12	**F:**CCAGCCTTCAACACCAAAGT **R:**CACGGGACATTCTACAAGCA	Std60°C	None	None	-
13	**F:**TGCTGTGTTTTATGTGGCGG **R:**CCTAGGTGAGTTTCGAGTGGT	Std58°C	1. p.Thr386Thr	2. p.Thr386Thr	1. 20/83 2. 15/55 (rs16935840)
14+15	**F:**TGTCCGCTGATCTTACCTTGT **R:**ACACAATGCTGCAATGAACATC	TD 65°C - 60°C	1. DIV: c.1488+36G>A c.489-41A>G	2. DIV: c.1488+36G>A c.489-41A>G	1. 34/83 2. 34/55
16	**F:**AGACCTAGCAATACCGACACA **R:**GCCCAGCCCAAATCAATAGG	Std58°C	None	N/A	-
17	**F:**GGTTTGGAATGAAATGTGCCA **R:**CCTAAAGGGCAGACTGACTATAA	Std60°C	None	None	-
18	**F:**CCAGATGGGCTTTAATGTCTCT **R:**AAGTGGTTGGCTA	Std60°C	None	None	-

	GAGAAATGT				
19	**F:**TCCTCACCTCCCCAAGAAAC **R:**AGTGAATACAAGTGCTTTCCAGA	-	N/A	N/A	-
20	**F:**AGTTATGAGAGTGGAGGCTTTT **R:**TGCCAACAACAAACCAGAAAA	TD 65°C - 60°C	1.p.Ser650Ser	2. p.Ser650Ser	1. 34/83 2. 12/55
21	**F:**AGGACGACAGAGTGAAACCC **R:**CATGAGCCACTGTGCCCA	Std58°C	None	None	-
22	**F:**TGGCCCATTTGTTGAAACCT **R:**CCAAATGGTGATGGCAATGATC	TD 65°C - 60°C	None	None	-

Table 11. A Summary of Variants Identified in CUL2 from the Cohort of 55 Tumour DNA and 83 Constitutional Blood DNA samples from familial and sporadic PCC/PGL patients. *Previously reported polymorphisms are outlined in purple, unreported synonymous variants are outlined in orange and unreported variants of unknown clinical significance are outlined in green (DIV refers to a deep intronic variant). It was not possible to complete sequencing for exon 19, and therefore this is marked with N/A.*

In a similar situation to *CUL3*, there were no particular variants of interest identified in *CUL2*; other than one variant of unknown clinical significance, that was predicted to have a benign effect on protein function by SIFT and PolyPhen. However, as previously,

it was not possible to obtain clean, confident sequencing for all samples on every plate with the time restrictions, and therefore, sequencing data for some exons was not obtained. Sequencing for exon 19 was not obtained as it was not possible to obtain sufficient amplification at the PCR stage. This is likely due to poor primer design; however, given there was insufficient time to investigate this further. The smallest combined sample size of familials and sporadic samples sequenced for an exon was 48. (A summary of sequencing can be found in supplementary table 3)

The variant of unknown significance identified (p.Lys109Glu) was predicted to have a benign effect on protein function by *in silico* software prediction tools, SIFT and PolyPhen (figure 28).

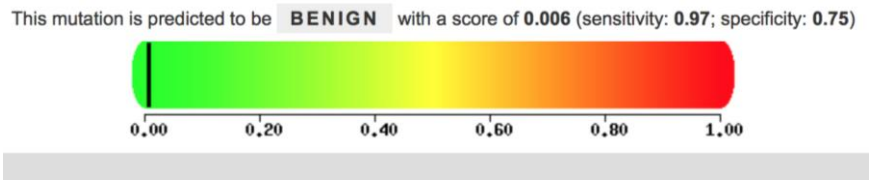


Figure 28. PolyPhen Prediction of p.Lys109Glu on CUL2 Function.

Although this work could potentially be continued by other members of Prof. Maher's team, the lack of potentially pathogenic alterations identified during this screening process have dampened the hypothesis that loss of function exonic alterations in *CUL2* are playing a contributory role in the development or progression of PCC/PGL tumours. Therefore, it is unlikely that investigations into *CUL2* will be continued.

5.3 Discussion of Sporadic Pheochromocytoma and Exome Sequencing

5.3a HIF-2α and Sporadic Pheochromocytoma and Paraganglioma

Unlike alterations in tumour suppressor genes, where numerous different mutations at various locations can result in truncated proteins with loss of function, cancer-related mutations occurring in oncogenes are often present in specific 'hotspot' regions that lead to a very specific activation or overexpression of the protein. Therefore, the initial identification of a variant in the exome sequencing data in *HIF2A* at a critical residue was highly interesting.

Exome sequencing data received for 15 tumour DNA samples (13 with apparent sporadic PCC/PGL) revealed the presence of two alterations in *HIF2A* (p.Pro531Thr, and p.Ser372Asn). The 531 residue corresponded to a primary hydroxylation site (as displayed in figure 29); this indicated that this alteration was highly likely to affect protein function. Furthermore, during this project, several publications became available

describing the identification of novel activating mutations in HIF-2α present in somatic cases of PCC/PGL. The studies identified several specific variants that resulted in increased HIF activity, including the same p.Pro531Thr variant found in the exome sequencing data of this study (Zhuang et al. 2012; Comino-Méndez et al. 2013, Lorenzo et al. 2013). Keeping in mind that alterations with pathogenic potential tend to cluster in critical proteomic regions, the finding of a variant located at a critical residue, that was also described to be pro-tumourigenic in other, reputable studies strongly indicated that the p.Pro531Thr alteration was likely to be acting as a driver in this particular tumour (Toledo *et al* 2013, Zhuang *et al* 2012).

Further mutational screening of *HIF2A* by Sanger sequencing of exons 9 and 12 enabled the identification of two further alterations (p.Pro407Arg and p.Phe583Leu). Interestingly, the p.Pro407Arg was also found to affect a residue in close proximity to another key hydroxylation site (Pro405), as illustrated in figure 29; and furthermore, it was found to exist in the same tumour that harboured the p.Pro531Thr alteration.

If these two variants (p.Pro531Thr and p.Pro407Arg) are present on different alleles, (*in-trans),* then one possibility is that the individual may be a compound heterozygote for *HIF2A*. It is also possible that one of the variants could have been inherited, which may have instigated genomic instability resulting in the coincidental acquirement of the second alteration. It could also be the case that both variants are present in a singular

tumoural clone and that their intraclonal redundancy is exerting a pro-tumourigenic effect through HIF-2α stabilisation. Alternatively, and the more likely possibility, given the low peak heights in figure 22 and figure 23, is that these two alterations are representative of two independent clonal populations, in which the co-operative effects of both may confer a growth advantage to the tumour; a particularly attractive suggestion, given that both of the variants occur at, or within close proximity to the two specific proline residues associated with hydroxylation.

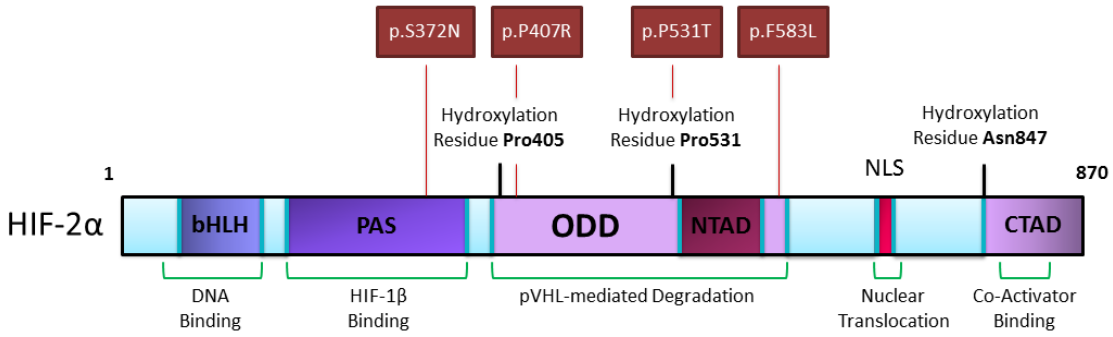


Figure 29. ***Protein Schematic of HIF-2α Including Location of Protein Domains adapted from (Rocha 2007).*** *This schematic illustrates the protein domains of HIF-2α, and annotates variants identified in this study. (Diagram not to scale)*

In *HIF2A*, as mentioned, codons 531 and 407 encode 2 critical residues that regulate HIF-2α stability through hydroxylation; therefore alterations at this point can have a substantial effect on protein function. Consistently, Toledo et al. 2013 showed that stable

expression of mutants at the 531 residue lead to increased stability of HIF-2α under normoxic conditions, caused by a conformational change to the hydroxylation sites in HIF-2α, which then prevented its recognition by PHDs. In addition, they showed that the specific substitution of Proline for Threonine in particular, lead to significantly larger tumour formation in xenograft mouse models in comparison to other substitutions at the same residue, suggesting that the specific alteration is particularly damaging (Toledo et al. 2013). It is therefore assumed that p.Pro532Thr is pathogenic, and is likely to have played an active role in the tumour in which it was identified in our study. With recent findings in the literature, it is assumed that the protumourigenic effects of this mutation prevent the association of HIF-2 α with PHDs and a consequential inability to be degraded by pVHL; this then leads to an accumulation of HIF-2α, ultimately resulting in the activation of a downstream signalling cascade mediated by the transcription of numerous HIF target genes (Kondo et al. 2002).

It would be interesting to know whether or not these variants were somatically acquired alterations or present in the germline; however, as no constitutional blood DNA was available for this patient, it was not possible to determine the somatic/germline status of these variants. However, in the electropherogram of the p.Pro531Thr alteration, it appears that the peak representing the variant allele is low; suggestive of the fact that this variant is more likely to be a somatically acquired alteration, which supports the initial hypothesis that the two independent clonal populations are the likely reason for

the presence of both variants. This would also provide an explanation as to why the p.Pro407Arg variant was not picked up initially in the exome sequencing data. For example, the tumour biopsy used to acquire the tumour DNA used in this study may have selected a small proportion of the tumour, and therefore there may have been a biased representation of the variants present. Therefore, the p.Pro407Arg variant may have been diluted amongst the more prominent cellular populations, and so it may have been called as a low quality read that was subsequently excluded during variant calling.

All of these hypotheses could be explored in future studies by investigating the downstream effects of both variants in known susceptibility pathways *in vitro* to determine the effect of harbouring two activating mutations located at these two specific proline residues by comparing them to cells separately harbouring each alteration alone.

The p.Ser372Asn alteration has not previously been reported, but based on its position, it is predicted to affect the PAS domain of the HIF-2α protein; the region in which HIF-1β binding occurs (figure 29). The significance of this alteration is uncertain, particularly as *in silico* predictions from SIFT and PolyPhen have suggested that the pathogenicity of this variant is likely to be benign. However, although this particular variant has not previously been described, a different variant, 2 nucleotides away, at residue 374 has previously been found to be associated with PCC/PGL in individuals with

polycythaemia (a condition associated with elevated levels of haemoglobin) (Lorenzo et al. 2013). The study of Lorenzo *et al* 2013, reported that they did not find any evidence of loss of heterozygosity or the presence of any additional somatic alterations, which they suggested was indicative that this variant played a predisposing role as opposed to a causative role in this disease.

Due to the close proximity of residues 372 and 374, it is possible that the tumourigenic outcomes observed by Lorenzo *et al* 2013 associated with the 374 residue may also indicate that alterations to 372 may have a similar tumourigenic outcome,

However, it should be noted that the 374 residue in *HIF2A* is highly conserved, suggesting a functionally important role; however the 372 residue is only moderately conserved (different in zebrafish and horse), and therefore it is also possible that regardless of their close proximity, this residue may be less functionally important (figure 30). All things considered, it would be necessary to carry out further investigations into the downstream effects of this variant *in vitro* in order to fully understand the tumourigenic implications of this variant.

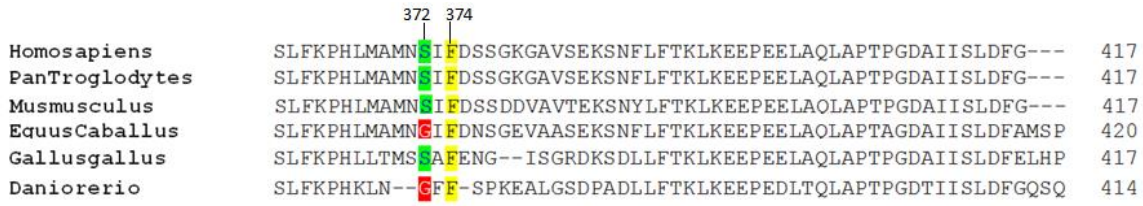


Figure 30. Multiple Sequence Alignment of HIF2A including residues 372 and 374

The patient found to carry the p.Ser372Asn alteration was also positive for a pathogenic *SDHB* alteration, which may suggest one of two options with regard to the HIF-2α alteration: the first possibility is that the inherited *SDHB* mutation resulted in the development of PCC/PGL and the *HIF2A* alteration identified in the tumour DNA was acquired and may be driving tumour progression. The second option is that both the p.Ser372Asn alteration and the pathogenic *SDHB* alteration (p.Trp200Asn) were inherited and that the cooperative effects of both variants contributed to tumour development. The corresponding normal blood DNA was unavailable for this individual, however, the electropherogram for this variant does seem to show an equal distribution between the variant and wildtype allele, which could be an indication that this alteration is present in the germline. It is interesting to consider the possibility that both may be germline; particularly as pathogenic HIF-2α variants have not previously been reported to occur in combination with other inherited PCC/PGL susceptibility genes; however, it will not be possible to know for sure without access to the corresponding normal blood DNA of this individual.

5.3b HRAS and Sporadic Pheochromocytoma and Paraganglioma

HRAS is one of the most commonly mutated genes in human cancer and disruption of the RAS/RAF/ERK and PI3K/AKT/mTOR kinase signalling pathways have been well documented to be associated with the development of PCC/PGL. However up until recently, mutations in *HRAS* had not previously been found to occur in PCC/PGL tumours. This may be due to the fact that many previous genetic studies into sporadic cases of PCC/PGL had largely focused on the presence of known susceptibility genes in tumour DNA. However, the recent publication by Crona *et al* 2013 demonstrated that somatic alterations in *HRAS* do indeed contribute to tumour progression, and as of yet germline alterations in *HRAS* have not been associated with the development of the inheritance of PCC/PGL. However, this is likely to be due to the fact that germline alterations in *HRAS* are associated with Costello syndrome; a condition belonging to a group of disorders known as the RASopathies (Gripp & Lin 2012). Interestingly, neither pheochromocytomas nor paragangliomas have been associated with Costello patients, which suggests that although the same pathway is disrupted, there must be additional genetic or environmental factors contributing to the non-penetrance of HRAS-associated PCC/PGL in these patients. It is likely that HRAS plays a specific, essential role during embryogenesis that becomes obsolete later in life, which is why germline and acquired HRAS alterations do not result in the same clinical phenotype.

Tumour DNA samples from 55 patients with apparently sporadic PCC/PGL were screened for mutation in *HRAS, KRAS* and *NRAS* by Sanger sequencing. These samples and a further 30 PCC/PGL DNA samples were also screened by Andrea Luchetti using the ampliseq IonTorrent cancer hotspot panel. No variants of interest were identified in *KRAS* or *NRAS*; however, our collaborative efforts revealed a total of 6 alterations from 85 patients in *HRAS* (1=p.Gly13Arg and 5=p.Gln61Arg). In total, this amounted to a total frequency of approximately 7.1%. In combination with mutations reported in the current literature, it is estimated that the mean frequency of *HRAS/BRAF* mutations in sporadic PCC/PGL is 8.9% and in PCC/PGL with an inherited gene mutation 0% (Luchetti et al. 2015; Oudijk et al. 2014; Crona et al. 2013). These figures suggest that *HRAS/BRAF* mutations and inherited PCC/PGL genes mutations might occur in a mutually exclusive manner (Luchetti et al. 2015).

Due to the fact that the mutations identified are known to affect two essential residues responsible for GTP-hydrolysis, our data in combination with current literature, indicates that mutations to these particular residues will likely result in the aberration of GTP-hydrolysis in HRAS resulting in the constitutive activation of the RAS pathway. This signals to a large downstream signalling cascade of various pathways including the RAS/RAF/ERK and PI3K/AKT/mTOR pathways which ultimately generates a complex web of pro-tumourigenic cellular consequences including increased cellular proliferation

and growth, RAS-induced apoptotic resistance and increased angiogenesis (Pylayeva-Gupta et al. 2011).

The identification of HRAS as a driver of tumourigenesis in PCC/PGL may have future clinical implications and could potentially highlight a novel therapeutic opportunity; although most common methods of treatment usually involve surgical removal of the tumour. The first descriptions of *BRAF* mutations in PCC/PGL may also contribute new perspectives on PCC/PGL progression; particularly as *BRAF* causes ERK kinase activation as part of the RAS/RAF/ERK signalling pathway, which indicates that constitutive RAS pathway activation can occur through mutation to various parts of the pathway. It would therefore be interesting to use a different targeted next generation sequencing panel containing more genes indicated in this pathway in order to determine whether any further somatic mutations can be identified in related genes in PCC/PGL tumours.

The collaborative finding of these mutations in *HRAS* associated with sporadic PCC/PGL highlighted the benefits of performing mutation screening using more than one technology. There are clear benefits and shortcomings to Sanger sequencing, whole exome sequencing and targeted next generation sequencing; however in this particular instance, the combination of these technologies was able to deliver positive results that would not likely have been achieved otherwise. Using the targeted next generation

sequencing panel to sequence the hotspot regions of known cancer genes meant that many patient samples could be sequenced in parallel with higher depth and greater coverage than would have been achieved with exome sequencing alone. This targeted method is also more sensitive to picking up mosaicism; a clear advantage over Sanger sequencing, where these types of variant can be missed. Using an NGS panel generates results much faster at a lower overall cost than would be achieved by Sanger sequencing as it sequences multiple reads in parallel. It seems reasonable to assume that genes already known to play a role in other cancers may also be likely to play a role in PCC/PGL tumours; however aside from the *HRAS, BRAF* and *TP53* alterations reported, no variants were identified in any of the other known cancer genes. This fact emphasises the genetic complexities of cancer, in that established connections between genes and aberrant molecular pathways in one tumour type cannot always be applied to others. For this reason, targeted next generation sequencing, although a powerful tool in its own right, is at its most useful when a regions and genes of interest have already been identified, and it is not often effective as a tool for novel gene discovery.

5.3d KEAP1 and CUL3 as Drivers of PCC/PGL Tumour Progression

All coding exons of *KEAP1* and *CUL3* were analysed by Sanger sequencing in a total of 55 tumour and 83 constitutional blood DNA samples from patients with suspected cases of sporadic or inherited PCC/PGL. Although one heterozygous alteration (c.1555A>G;

p.Ile519Val) was identified in the tumour DNA of a patient with sporadic paraganglioma, the variant was not predicted to affect protein function. There are several possibilities that could explain the finding of a single alteration from a panel of 138 patients. For example, the variant could simply have been inherited as a rare, benign polymorphism. Alternatively, the variant could also have occurred as a sporadic event in the tumour. The possibility that this variant could be playing a tumourigenic role in the progression of a minute proportion of PCC/PGL patients (<1%) can also not be excluded.

CUL3 was investigated toward the late stages of this project, and therefore as a result of time restrictions to the project, some exons could not be sequenced successfully; however, the minimum number of successful patients in an incomplete exon was 20. Although not ideal, a minimum cohort of 20 was still considered to be a reasonable number of patients, which would enable the detection of higher frequency variants, although perhaps not those at low frequency. Unfortunately, exons 11 and 13 were not able to be sequenced as no PCR products could be obtained.

KEAP1 and *CUL3* were not investigated for the presence of epigenetic alterations or for the presence of larger exonic duplications or deletions, and so the possibility that KEAP1 or CUL3 protein function is potentially disrupted by these factors cannot be ruled out. However, the lack of further intragenic alterations in either of these genes in

the total cohort of 138 patients meant that it was not viable to continue investigations into *KEAP1*or *CUL3* as potential drivers of PCC/PGL

Recent studies have found relatively low levels of exonic mutations in *CUL3* and *KEAP1*, but have found high numbers of tumours with copy number loss and hypermethylation of the promoter regions of both *KEAP*1 and *CUL3* (Martinez et al. 2014; Sporn & Liby 2012). Inactivating mutations affecting *KEAP1* and *CUL3* have been identified in a wide variety of tumour types including ovarian, thyroid, head & neck and non-small cell lung tumours (Martinez et al. 2014). Furthermore, Martinez *et al* 2014 also showed a marked enrichment of Nrf2 target genes in ovarian tumours harbouring disruptions to KEAP1-CUL3 complex; this highlights the likely effects of Nrf2 pathway hyperactivation and is suggestive that this is contributing to tumour progression in these cases. Given these findings in other tumours, it would be interesting to investigate the expression levels of KEAP1, CUL3 and Nrf2 in PCC/PGL tumours in order to determine if activation of this pathway correlates with disease severity; particularly in FH-deficient PCC/PGL tumours as this was found to be the mechanism of tumourigenesis in FH-deficient renal tumours (Adam et al. 2011).

Given the findings in this study in combination with the information available in current literature, it cannot be said with certainty whether or not disruptions in KEAP1 or CUL3 are playing a role in the development or progression of PCC/PGL; however it is

hypothesized that copy number loss and hypermethylation may potentially be identified if KEAP1 and CUL3 are investigated in a larger sample size of PCC/PGL patients.

5.3e Cullin-2 and PCC/PGL

CUL2 was screened in a cohort of 55 tumour DNA and 83 constitutional blood DNA samples from individuals with sporadic and familial forms of PCC/PGL. Although not every exon was sequenced to completion in the entire cohort of patients, the lowest number of patients sequenced in an exon was 48; which was considered to be acceptable given the lack of time. The lack of potentially pathogenic alterations found from those sequenced meant it was not economically or temporally feasible to continue to complete the sequencing in these individuals. Furthermore, shortly after completing the first stages of the mutation screening process, some publications were released which had also investigated *CUL2* as a potential driver of PCC/PGL tumourigenesis. It was therefore decided that a review of the current findings in this project and a comprehensive review of the literature should be performed in order to determine the feasibility and practicality of continued investigations into *CUL2*.

The variant of unknown significance that was identified (p.Lys109Glu) has been investigated using *in silico* mutation prediction software tools in order to generate a rough prediction of its deleterious potential. *In silico* tools, SIFT and PolyPhen

suggested that this variant of unknown significance was predicted to have a benign effect on protein function. Although *in silico* tools are not completely reliable, they are useful for supporting theories of pathogenicity. They use a combination of factors to produce the predicted score, including levels of conservation of the amino acid across species and the location of the amino acid in relation to the essential functional domains of the protein. In these cases for *CUL2*, although it cannot be said with full confidence, it is predicted that (p.Lys109Glu) is unlikely to have a major effect on protein function, and so it was decided not to pursue investigations into this gene any further.

Other deciding factors for terminating investigations into the potential roles of the genetic inactivation of CUL2 in PCC/PGL tumourigenesis, included the fact that a relevant publication was recently released during the *CUL2* mutation screening process. Rowbotham *et al* 2014 also investigated disruptions to various parts of the hypoxia pathway in order to determine potential involvement in the progression of PCC/PGL tumours. They found that in a panel of 171 PCC tumours, approximately 60% were found to harbour copy number loss alterations to at least one component of the hypoxia pathway, and that these alterations correlated with increased expression of hypoxia target genes (Rowbotham et al. 2014). Surprisingly, and of relevance to this study, they did not find any form of genetic mutation, copy number loss or hypermethylation to occur in *CUL2*. It was suggested that perhaps epigenetic inactivation of *CUL2* may be

playing a role in tumour development; however without comprehensive expression studies of *CUL2* in PCC/PGL tumours, it is not possible to know for sure.

In conclusion, *CUL2* did initially seem a reasonable candidate gene for involvement in the progression or development of PCC/PGL tumours, given its association with VHL and its involvement in the hypoxia pathway. The fact that other labs were simultaneously investigating the role of *CUL2* in PCC/PGL tumours supports our initial theories of its associations with these tumours. Unfortunately, although other researchers have identified different components of the hypoxia pathway to be associated with PCC/PGL tumours, multiple parts to their investigations have discounted the likelihood of the involvement of CUL2. The fact that no clearly pathogenic variants were identified in our cohort of 138 PCC/PGL patients in combination with the lack of alterations identified in the recent literature from a further 171 patients is strongly suggestive that genetic inactivation of CUL2 does not play a role in the pathogenesis or progression of PCC/PGL.

Therefore, taking into account the lack of clearly pathogenic variants found in *CUL2* in a total number of 309 PCC/PGL patients (171 samples from the literature, and 55 tumours and 83 constitutional blood DNA samples from this study) it is hypothesised that inactivation of *CUL2* by genetic mutation is unlikely to play a role in the pathogenesis or progression of these tumours.

5.3f Sporadic PCC/PGL and Exome Sequencing

There are clearly many potential benefits to be gained from exome sequencing projects investigating the genetic contributions to the development and progression of somatic tumours. Successful identification of novel genes and variants that are found to play an active role in these tumours may uncover valuable, previously unknown, information regarding tumour biology. These findings may also shed light on novel mechanisms of tumour progression, which may then potentially lead to opportunities for therapeutic intervention. These successes have been achieved on many occasions by various different studies.

It should be noted that along with the potential benefits of these types of studies also come vast challenges and difficulties, which must be overcome. The most prominent challenges faced during this study included the filtration and prioritization of the vast number of candidate variants in addition to the limitations of manpower with regard to carrying out mutational screening. Although no novel genes were identified to play a role in the progression of somatic PCC/PGL tumours, this study did uncover several beneficial discoveries. Firstly, this study identified novel alterations occurring in both *HIF-2α* and *HRAS*; both of these discoveries may be of clinical relevance in the near future with regard to diagnostics and even possibly therapeutic choices. The use of the IonTorrent cancer hotspot panel facilitated the successful identification of these novel alterations in *HRAS*, and also provided the data necessary to observe that alterations in

HRAS segregated with somatic tumours. This discovery in particular may benefit patients in the future, not only to assist diagnostics and guide treatment plans, but may also provide *HRAS* positive PCC/PGL patients with the valuable information and peace of mind that their family members are unlikely to be at an increased risk of developing these tumours.

In conclusion, although these projects have their challenges, the potential for them to produce results of clear clinical benefit is unmistakable. With continued investigations into the predisposition, development and progression of these tumours, it is hoped that effective preventative and therapeutic treatments may be developed to help and support those affected.

Chapter Six:
DISCUSSION

6.1 Summary of Findings

In a brief summary of all findings, exome sequencing was employed as an investigative tool in an attempt to identify genes associated with a range of different disorders, including rare, autosomal recessive disease and both familial and sporadic cancer. Due to the different genetic natures and complexities between these disorders, there was quite a significant variation in the degrees of success of each section of the project. These will be summarised briefly in the following sections.

6.1a Evaluation of Exome Sequencing for use in Recessive Disease-Gene Discovery

Exome sequencing was used to screen one individual from a consanguineous family who was affected with Acrocallosal syndrome in the hope of identifying a causal variant responsible for their disease. Exome sequencing was used as it initially enables an unbiased approach to gene discovery. The fact that acrocallosal syndrome is an autosomal recessively inherited disorder was of great benefit, as it allowed the removal of all heterozygous alterations, and so only rare homozygous alterations remained. This reduced the number of candidates from 28,588 to 9; a substantial reduction of 99.97% of variants. This process was facilitated further by the fact that DNA from multiple additional family members was available, thus providing the opportunity to carry out segregation studies, thus enabling the confident exclusion of non-segregating variants. Consanguineous families, as mentioned previously, will share many of their alterations,

and therefore offer a powerful means of gene detection. This in combination with a rare, autosomal recessively inherited disorder meant that the odds of identifying the disease gene were quite favourable. Consistently with this, a homozygous deletion (c.653_662delCACGGTCTT, p.(His218_Phe221delinsLeu)) was detected in *KIF7* in the affected individual, and was successfully shown to segregate with the disease status within the family.

A publication was released at the time of this discovery, describing mutations in *KIF7* associated with the inheritance of acrocallosal syndrome (Putoux et al. 2012). However, interestingly, it was suggested that the presence of modifier genes may be playing a role in the wide array of phenotypic variability that is often observed in this condition. As a result of this, our continued investigations managed to identify the presence of alterations in *AHI1:* c.C2488T p.Arg830Trp; *BBS2*: c.G209A p.Ser70Asn and *BBS4* c.A1414G p.Met472Val; variants which have all previously been shown to be present in other ciliopathies. Whether these alterations are truly contributing a modifying effect to the phenotype observed in the affected individual from this study will remain speculative until more comprehensive analytical studies have been performed in larger cohorts of affected individuals.

These findings demonstrated the power of exome sequencing to successfully identify disease variants in autosomal recessive disease. This data was published in the European

Journal of Medical Genetics in 2013; and it is hoped that these findings will be of some clinical benefit to patients affected with ciliopathies such as acrocallosal disease.

6.1b Evalutation of Challenges and Successes of Exome Sequencing for use in Disease-Gene Discovery for Inherited Cancer

Following the successful gene identification with autosomal recessive disorders, it was considered that the powerful abilities of next generation sequencing technology could be used to investigate the genetic causes of inherited cancers. Pheochromocytoma is one of the most genetically heterogeneous cancers, and the number of genes conferring increased susceptibility to these tumours is constantly growing. As many tumours with a strong pattern of heritability do not harbour a mutation in any of the known susceptibility genes, it is likely that there are many genes yet to be discovered.

Familial cancers are often inherited in an autosomal dominant pattern with incomplete penetrance, which can add a certain complexity to gene identification studies. In addition, the cancer genome has a highly genetically heterogeneous landscape that can be influenced by multitude of factors including environmental exposures. These highly variable aspects to these disorders mean that employing exome sequencing to carry out gene discovery investigations can be extremely challenging in comparison to studies in autosomal recessive disorders. Firstly, the autosomal dominant inheritance patterns mean that no heterozygous variants can be excluded from the data, which greatly

increases the number of potential candidates. Furthermore, the incomplete penetrance in cancer means that variants cannot be excluded based on their presence in unaffected individuals alone. The highly diverse and complex mechanisms of tumourigenesis also means that multiple different molecular pathways containing vast numbers of genes may potentially be playing a tumourigenic role, which again, expands the number of possible candidates exponentially. Nevertheless, our approach was to compile groups of DNA samples from a number of affected individuals, who had clear indications of an inherited form of PCC/PGL, but who did not carry a germline mutation in any susceptibility gene. In the absence of available DNA from comprehensive families with PCC/PGL patients, our approach was to attempt to identify certain genetic similarities between unrelated, affected individuals in the hope that disease variants would be enriched in the disease population in comparison to healthy controls.

Exome sequencing was therefore performed for 11 unrelated PCC/PGL patients which resulted in a collective quantity of 263,798 variants. Even following the vigorous and highly stringent variant filtration methods applied to reduce this number, there were still a potential 3809 variants (766 of which were either frameshifts, nonsense or splicing). At this point, it was necessary to apply a candidate gene approach; however, many of the genes present were uncharacterised and therefore, without any functional evidence, could not be investigated. A further challenge to this study was the fact that heterozygous variants are associated with a higher frequency of false positive results in comparison to homozygous variants. This is because 100% of the reads should map to

the altered allele for a homozygous variants, while heterozygous variants are usually called when 50% of the reads map to the altered allele. For this reason, during the inevitable errors that are encountered during the sequencing process, it is easier for false positive heterozygous variants to be called. This can be countered by performing deeper sequencing so that that the quantity of false reads may be diluted. Even with sufficient read depths, averaging between 30-100, numerous false positives were encountered (data not shown). A higher frequency of false positives were encountered for missense variants, as statistically, a single base change that results in an amino acid substitution is likely to be more common than a specific single base change that results in a nonsense. Frameshifts were the least likely to be false positives. These factors in addition to the deleterious potential of these variants supported our main strategy to primarily focus on nonsense, frameshift and splicing variants, which collectively amounted to 766 variants. These 766 variants were then scrutinized in order to determine any potential functional relevance with related cancer pathways, which resulted in the selection of *UBE2C, ERF,* and *ME2*.

Comprehensive mutational analysis of these genes in a panel of 77 familial pheochromocytoma samples did not reveal any variants of interest apart from several variants of unknown significance (that were also detected to be benign by *in silico* mutation prediction tools).

The lack of variants could be due to many reasons, and does not necessarily signify that these genes are not contributing to PCC/PGL progression in some cases. For example, many studies that have successfully identified novel PCC/PGL susceptibility genes have reported mutation frequencies of less than 1.5% of patients (i.e *MAX*: 1.12% of patients, *FH*: 0.83% of patients and *TMEM127*: 0.9% patients) (Castro-Vega et al. 2014; Comino-Méndez et al. 2011; Qin et al. 2010). These studies managed to successfully identify these genes either by including large sample sizes (598 patients were screened to discover *FH*), or through the use of large families containing multiple affected individuals, followed by segregation studies. For these reasons, a cohort of 77 patients, meant that it was unlikely that a sufficient number of patients would have been identified in order to successfully establish a statistically significant connection between PCC/PGL susceptibility and low frequency variants (present in <1% of patients).

The possibility that any of these genes contain large duplications or deletions, or harbour alterations to promoter/enhancer regions has not been explored. It is also possible therefore that these genes are altered in PCC/PGL tumours, but perhaps not by intragenic mutation. In order to confidently determine that these genes are not playing a PCC/PGL-related tumourigenic role, larger studies in larger cohorts of patients (ideally large families) would need to be performed to rule out the possibilities of low frequency intragenic mutations. In addition, studies in these larger cohorts would need to be carried out using MLPA to rule out the possibility of duplications or deletions which may not

have been picked up through Sanger sequencing. Investigations into the relevant levels of protein expression in PCC/PGL tumours for ERF, ME2, UBE2C and UBE2QL1 would provide a further indication as to whether these proteins are aberrant in these cancers. Finally, methylation and loss of heterozygosity (LOH) studies could also have been performed to determine whether the second allele had been lost in the tumours (which would have supported a potential tumourigenic role). Many further studies could have been performed; however, as there was not a substantial amount of evidence available, it was not justifiable to carry out these investigations for this particular study.

6.1c Evaluation of Exome Sequencing as a Tool to Investigate Drivers of Sporadic Cancer

Following investigations into familial cases of PCC/PGL, it was decided that exome sequencing technology would be used in an attempt to identify potential drivers of tumourigenesis in sporadic cases of PCC/PGL. It was thought that there would be a greater possibility of identifying tumour-associated variants by sequencing tumour DNA, particularly as there are likely to be multiple variants driving tumour progression, whereas in inherited cases, it is likely that only a single germline mutation was involved in conferring increased susceptibility.

The approach was therefore to sequence DNA from multiple tumours in order to determine whether similar patterns of variants could be identified across multiple tumours. As most PCC/PGL tumours fall into a distinct molecular profile (either

pseudohypoxic or enriched kinase signalling), it was decided that genes present in these pathways would be prioritised.

A similar series of challenges were encountered during the gene prioritisation process as when this was performed for the familial PCC/PGL exome sequencing data. Even after rigorous variant filtration methods, the number of candidate variants was reduced from 394,673 to 3884; however this was still an extremely large number. Studies into sporadic cancer aim to identify genes that have occurred somatically, and therefore these alterations should be present in the tumour DNA, but absent from the normal tissue or corresponding constitutional blood DNA. It is therefore a useful technique to sequence corresponding normal DNA along with the tumour DNA so that variants that are common to both samples can be excluded. This can remove vast numbers of candidates, allowing the researcher to focus solely on somatic variants. However, this was not a viable option for this project, as there was no normal DNA available for many of the samples sequenced. It was therefore considered most efficient to sequence DNA from a larger number of tumours.

Several publications that were released during the course of this study instigated investigations into *HIF2A* and *HRAS;* both of these genes were found to be mutated in the exome sequencing data (p.Pro531Thr and p.Ser372Asn) in *HIF2A* and (p.Gly61Arg) in *HRAS*. It cannot be known as to whether these variants would have been selected as

candidate genes in this study; however, it is possible that *HIF2A* would have passed through many of the selection criteria applied to this study. There were two different variants present in two different tumours, the variants were absent from SNP databases, and furthermore the functional relevance of the gene in hypoxic pathways was very appropriate. However, the variant in *HRAS* is more questionable, as only one missense variant was identified in the exome sequencing data; in this case, there were a mammoth number of other genes that also fit this criteria. Nevertheless, the findings of other groups can instigate multiple additional studies, which can then discover further information that can also be of clinical benefit. Thus, our further investigations into these genes uncovered novel, unreported variants (p.Pro407Arg and p.Phe583Leu in *HIF2A* and p.Gln61Arg in *HRAS)* that may be contributing to tumour progression.

The application of a targeted next generation sequencing approach using a cancer gene panel in the cohort of sporadic PCC/PGL patients proved to be a very useful technique for identifying variants rapidly, at low cost in large number of individuals. Furthermore, this technique was more sensitive to mosaic alterations, and so, it was able to identify a further 2 (p.Gln61Arg) alterations in *HRAS* that had previously been missed by Sanger sequencing, in addition to 1 variant in BRAF and 2 variants in TP53 (genes which had not been sequenced by Sanger). This approach can be very effective, particularly for use in routine diagnostics due to the large numbers of genes covered, high depth and accuracy. However, this technique is not as effective when attempting to carry out gene

discovery (highlighted by the fact that no other alterations were observed in any of the genes on this panel).

6.1c.i Clinical Relevance of Findings in HIF2A and HRAS

These findings have expanded upon the literature and have increased current understandings of the PCC/PGL tumourigenic landscape. For example, *HRAS* mutations are only found in sporadic cases of PCC/PGL (as germline mutations are associated with Costello syndrome). Furthermore, *HRAS* mutations and mutations in known susceptibility genes currently appear to occur in a mutually exclusive manner. By observing multiple PCC/PGL patients carrying these mutations, associations may be made between these mutations and common tumourigenic characteristics, for example, patterns in the risks of malignancy or recurrent disease may be highlighted. This would then potentially enable patients to have access to a more comprehensive and informed prognosis. Furthermore, the finding of mutual exclusivity between *HRAS* and inherited mutations can provide clear indications on risks of inherited disease for additional family members. These findings in *HRAS*, as a collaborative effort with Andrea Luchetti, were published in the International Journal of Endocrinology in August, 2014. The collective use of exome sequencing, the IonTorrent targeted cancer panel, and Sanger sequencing facilitated the identification of multiple *HRAS* variants by enabling the massively parallel sequencing of many individuals with deep coverage. This

combinational approach highlights the benefits of using multiple methods of analysis to investigate genetic disease.

Interestingly, the findings from the *HIF2A* screening indicated two tumours, each containing two variants of interest; one tumour containing a pathogenic *SDHB* mutation (p.Trp200Asn) and a variant in *HIF2A* (p.Ser372Asn), and a second tumour containing (p.Pro531Thr) and (p.Pro407Arg) both in *HIF2A.* Unfortunately, the somatic/germline status of these alterations could not be identified in the absence of any constitutional blood DNA; however, as discussed earlier, it is likely that these variants represent independent clonal populations within the tumour. This information may provide a biological insight into the mechanisms of tumour progression in these individuals. Furthermore, and of clinical relevance, the (p.Pro407Arg) variant was present in an individual who had exome sequencing performed on their tumour DNA; however the (p.Pro407Arg) variant was not picked up. This could indicate that the tumour biopsy performed contained a biased representation of the variants present. This also highlights an important issue that may be encountered in the clinical environment when attempting to genetically characterise tumours for treatment choices. It is important to consider, as genetic intratumoural diversity may present a hindrance to personalised therapeutic approaches. Continued investigations that aim to characterise commonly mutated genes in these tumours may enable the development of specific combinational therapies that can be applied to target multiple aberrant pathways so that these situations can be

addressed clinically. It is hoped that the findings from this study may be of benefit in combination with future studies that collectively can translate into the clinical environment to improve preventative, diagnostic, prognostic and therapeutic strategies in the management of PCC/PGL patients.

6.1c.ii Evaluation of the Candidate Gene Approach

The candidate gene approach is sometimes criticised for its lack of reproducibility and its potential to incorporate subjectivity and bias. However, it is often difficult to avoid this when carrying out projects such as this. For this reason, the use of multiple techniques can dilute the effects of bias; for example, use of multiple different filtration techniques, use of bioinformatics tools, and incorporating other technologies such as the targeted cancer gene panel. Although limited success was seen following the screening of *KEAP1*, *CUL3* and *CUL2*, this does not necessarily mean that poor choices of candidates were selected; particularly as other reputable groups had also simultaneously been carrying out studies into the associations between these genes and PCC/PGL (Martinez et al 2014, Rowbotham et al, 2014). This not only suggests that the gene selections were within reason, but also that the lack of intragenic mutations identified in their studies also supports the absence of pathogenic intragenic mutations from this study.

6.1d Final Comments on Exome Sequencing to Investigate Disease

In conclusion, there can be numerous potential benefits that can be gained from exome sequencing projects; as demonstrated by the successes achieved with autosomal recessive disease and the relative successes from identifying novel mutations in *HIF2A* and *HRAS*; however these projects can also come with a substantial number of challenges. These projects are highly dependent on resources; for example, exome sequencing for use in the investigations of familial disease can be most successful in the presence of large comprehensive families containing multiple affected individuals. Exome sequencing can also be a powerful tool to investigate sporadic disease in the presence of large quantities of affected individuals displaying similar characteristics. However, it is not always possible to have access to such valuable groups of people. Nevertheless, as genetic investigations are continued, and novel discoveries continue to stem further projects, the genetic landscape of such diseases is becoming increasingly clear.

With the advent of new ventures such as the 100,000 Genomes Project, the genetic contributions to disease will only become clearer. It is hoped that in the near future, as costs continue to fall and throughput continues to rise, technologies such as exome sequencing will become even more accessible within the research environment, so that projects such as the one completed here, may continue to identify novel genes and pathways that can be targeted therapeutically to improve the lives of patients. It is

predicted that next generation sequencing may not only continue to provide novel informative discoveries, but may also become commonplace in the clinical environment, both as an investigative tool and also as a tool that can be used to efficiently predict and diagnose, and treat disease.

REFERENCES

Abu-Safieh, L. et al., 2012. In search of triallelism in Bardet–Biedl syndrome. *European Journal of Human Genetics*, 20(4), pp.420–427.

Adam, J. et al., 2011. Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling. *Cancer cell*, 20(4), pp.524–537.

Alkuraya, F.S., 2013. Impact of new genomic tools on the practice of clinical genetics in consanguineous populations: the Saudi experience. *Clinical genetics*, 84(3), pp.203–208.

Antoniou, AC, Pharoah, PP, Smith, P, Easton DF. 2004, The BOADICEA model of Genetic Susceptibility to Breast and Ovarian cancer. *Br J Cancer*, Oct 18:91(8):1580-90.

Astuti, D. et al., 2001. Gene mutations in the succinate dehydrogenase subunit SDHB cause susceptibility to familial pheochromocytoma and to familial paraganglioma. *American journal of human genetics*, 69(1), pp.49–54.

Auer, P.L. & Lettre, G., 2015. Rare variant association studies: considerations, challenges and opportunities. *Genome medicine*, 7(1), p.16.

Badano, J.L., Leitch, C.C., et al., 2006. Dissection of epistasis in oligogenic Bardet-Biedl syndrome. *Nature*, 439(7074), pp.326–330.

Badano, J.L., Mitsuma, N., et al., 2006. The ciliopathies: an emerging class of human genetic disorders. *Annual review of genomics and human genetics*, 7, pp.125–148.

Baker, K. & Beales, P.L., 2009. Making sense of cilia in disease: the human ciliopathies. *American journal of medical genetics. Part C, Seminars in medical genetics*, 151C(4), pp.281–95.

Bao, R. et al., 2014. Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. *Libertas Academica*, 13, pp.67–82.

Bardella, C. et al., 2012. Cells Lacking the Fumarase Tumor Suppressor Are Protected from Apoptosis through a Hypoxia-Inducible Factor-Independent, AMPK-Dependent Mechanism. *Molecular and Cellular Biology*, 32(15), pp.3081–3094.

Bayés, M., Heath, S. & Gut, I.G., 2012. Applications of Second Generation Sequencing Technologies in Complex Disorders. *Current topics in behavioral neurosciences*.

Baysal, B.E. et al., 2000. Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. *Science (New York, N.Y.)*, 287(5454), pp.848–851.

Bentley, D.R. et al., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), pp.53–59.

Biesecker, L.G. & Spinner, N.B., 2013. A genomic view of mosaicism and human disease. *Nature reviews. Genetics*, 14(5), pp.307–20.

Bishop, M., 1989. Retroviruses and oncogenes. *British journal of rheumatology*, 28(4), pp.350–357.

Bittles, A., 2001. Consanguinity and its relevance to clinical genetics. *Clinical genetics*, 60(2), pp.89–98.

Bittles, A.H. & Black, M.L., 2010. Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 107 Suppl, pp.1779–1786.

Blackwood, E.M., Luscher, B. & Eisenman, R.N., 1992. Myc and Max associate in vivo. *Genes {&} Development*, 6(1), pp.71–80.

Bourgeron, T. et al., 1994. Mutation of the fumarase gene in two siblings with

progressive encephalopathy and fumarase deficiency. *The Journal of clinical investigation*, 93(6), pp.2514–2518.

Boveri, T., 1914. *In Zur Frage der Entstehung Maligner Tumoren*, G. Fischer.

Burgunder, J.-M., 2014. Genetics of Huntington's disease and related disorders. *Drug discovery today*, 19(7), pp.985–989.

Burnichon, N. et al., 2011. Integrative genomic analysis reveals somatic mutations in pheochromocytoma and paraganglioma. *Human molecular genetics*, 20(20), pp.3974–3985.

Burnichon, N. et al., 2010. SDHA is a tumor suppressor gene causing paraganglioma. *Human molecular genetics*, 19(15), pp.3011–3020.

Burnichon, N. et al., 2012. Somatic NF1 inactivation is a frequent event in sporadic pheochromocytoma. *Human molecular genetics*, 21(26), pp.5397–5405.

Burrell, R. a et al., 2013. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501(7467), pp.338–45.

Castro-Vega, L.J. et al., 2014. Germline mutations in FH confer predisposition to malignant pheochromocytomas and paragangliomas. *Human molecular genetics*, 23(9), pp.2440–2446.

Cavalli, R., Varella-García, M. & Liang, C., 1997. Diminished Tumorigenic of Mitochondrial Phenotype after Depletion. *Cell Growth Differ*, 8(November), pp.1189–1198.

Chang, G.G. & Tong, L., 2003. Structure and Function of Malic Enzymes, A New Class of Oxidative Decarboxylases. *Biochemistry*, 42(44), pp.12721–12733.

Chang, Y.-L. et al., 2014. Human Mitochondrial NAD(P)+–Dependent Malic Enzyme Participates in Cutaneous Melanoma Progression and Invasion. *Journal of Investigative Dermatology*, 135(3), pp.807–815.

Chen, Z. et al., 2015. Cullin family proteins and tumorigenesis: genetic association and molecular mechanisms. *Journal of Cancer*, 6(3), pp.233–242.

Clark, G.R. et al., 2014. Germline FH Mutations Presenting With Pheochromocytoma. *The Journal of Clinical Endocrinology & Metabolism*, 99(10), pp.E2046–E2050.

Comino-Méndez, I. et al., 2011. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nature genetics*, 43(7), pp.663–7.

Comino-Méndez, I. et al., 2013. Tumoral EPAS1 (HIF2A) mutations explain sporadic pheochromocytoma and paraganglioma in the absence of erythrocytosis. *Human molecular genetics*, 22(11), pp.2169–2176.

Consortium, T.G. of the N., 2014. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature genetics*, 46(8), pp.818–825.

Cooper, D.N. et al., 2013. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Human genetics*, 132(10), pp.1077–1130.

Cooper, G.M. & Shendure, J., 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature reviews. Genetics*, 12(9), pp.628–640.

Crona, J. et al., 2015. Bioinformatic Challenges in Clinical Diagnostic Application of Targeted Next Generation Sequencing: Experience from Pheochromocytoma. *PloS one*, 10(7), p.e0133210.

Crona, J. et al., 2013. Somatic mutations in H-RAS in sporadic pheochromocytoma and paraganglioma identified by exome sequencing. *The Journal of clinical endocrinology and metabolism*, 98(7), pp.E1266–71.

Cullinan, S.B. et al., 2004. The Keap1-BTB protein is an adaptor that bridges Nrf2 to a Cul3-based E3 ligase: oxidative stress sensing by a Cul3-Keap1 ligase. *Molecular*

and cellular biology, 24(19), pp.8477–8486.

Curtis, V.F. et al., 2014. Stabilization of HIF through inhibition of Cullin-2 neddylation is protective in mucosal inflammatory responses. *The FASEB Journal*, 29(1), pp.208–215.

Cutting, G.R., 2014. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nature Reviews Genetics*, 16(1), pp.45–56.

Dafinger, C. et al., 2011. Mutations in KIF7 link Joubert syndrome with Sonic Hedgehog signaling and microtubule dynamics. *The Journal of clinical investigation*, 121(7), pp.2662–7.

Dahia, P.L.M., 2014. Pheochromocytoma and paraganglioma pathogenesis: learning from genetic heterogeneity. *Nature reviews. Cancer*, 14(2), pp.108–19.

Díaz-Flores, L. et al., 2008. Histogenesis and morphofunctional characteristics of chromaffin cells. *Acta physiologica (Oxford, England)*, 192(2), pp.145–63.

Dimova, N. V et al., 2012. APC/C-mediated multiple monoubiquitylation provides an alternative degradation signal for cyclin B1. *Nature cell biology*, 14(2), pp.168–176.

Donis-Keller, H. et al., 1993. Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. *Human molecular genetics*, 2(7), pp.851–856.

Gaude, E. & Frezza, C., 2014. Defects in mitochondrial metabolism and cancer. *Cancer {&} metabolism*, 2, p.10.

Gelman-Kohan, Z. et al., 1991. Further delineation of the acrocallosal syndrome. *European journal of pediatrics*, 150(11), pp.797–9.

Gerlinger, M. et al., 2012. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *The New England journal of medicine*, 366(10), pp.1107–1116.

Gossage, L., Eisen, T. & Maher, E.R., 2015. VHL, the story of a tumour suppressor

gene. *Nature reviews. Cancer*, 15(1), pp.55–64.

Gottlieb, E. & Tomlinson, I.P.M., 2005. Mitochondrial tumour suppressors: a genetic and biochemical update. *Nature reviews. Cancer*, 5(11), pp.857–866.

Greenman, C. et al., 2007. Europe PMC Funders Group Patterns of somatic mutation in human cancer genomes. *Nature*, 446(March), pp.153–158.

Gripp, K.W. & Lin, A.E., 2012. Costello syndrome: a Ras/mitogen activated protein kinase pathway syndrome (rasopathy) resulting from HRAS germline mutations. *Genetics in medicine : official journal of the American College of Medical Genetics*, 14(3), pp.285–292.

Hamamy, H., 2012. Consanguineous marriages : Preconception consultation in primary health care settings. *Journal of community genetics*, 3(3), pp.185–192.

Hanahan, D. & Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell*, 144(5), pp.646–674.

Hao, H. et al., 2009. SDH5 , a Gene Required for Flavination. *Science*, 325, pp.1139–1142.

Hardie, D.G., 2011. AMP-activated protein kinase: an energy sensor that regulates all aspects of cell function. *Genes {&} development*, 25(18), pp.1895–1908.

Hast, B.E. et al., 2014. Cancer-derived mutations in KEAP1 impair NRF2 degradation but not ubiquitination. *Cancer research*, 74(3), pp.808–817.

He, M. et al., 2014. The kinesin-4 protein Kif7 regulates mammalian Hedgehog signalling by organizing the cilium tip compartment. *Nature cell biology*, 16(7), pp.663–672.

Hershko, A. & Ciechanover, A., 1998. The ubiquitin system. *Annual review of biochemistry*, 67, pp.425–479.

Hichri, H. et al., 2005. Testing for triallelism: analysis of six BBS genes in a Bardet-

Biedl syndrome family cohort. *European journal of human genetics : EJHG*, 13(5), pp.607–616.

Hodgson, B.D., Davies, L. & Gonzalez, C.D., 2009. Acrocallosal syndrome: a case report and literature survey. *Journal of dentistry for children (Chicago, Ill.)*, 76(2), pp.170–7.

Hoefele, J. et al., 2007. Evidence of oligogenic inheritance in nephronophthisis. *Journal of the American Society of Nephrology : JASN*, 18(10), pp.2789–2795.

Illumina, 2015. An Introduction to Next-Generation Sequencing Technology.

Ingham, P.W. & McMahon, A.P., 2009. Hedgehog signalling: Kif7 is not that fishy after all. *Current biology : CB*, 19(17), pp.R729–31.

Jacks, T. et al., 1994. Tumour predisposition in mice heterozygous for a targeted mutation in Nf1. *Nat Genet*, 7(3), pp.353–361.

Jafri, M. & Maher, E.R., 2012. The genetics of phaeochromocytoma: using clinical features to guide genetic testing. *European journal of endocrinology / European Federation of Endocrine Societies*, 166(2), pp.151–158.

Jochmanová, I. et al., 2014. HIF signaling pathway in pheochromocytoma and other neuroendocrine tumors. *Physiological research / Academia Scientiarum Bohemoslovaca*, 63 Suppl 2, pp.S251–62.

Jochmanová, I. et al., 2013. Hypoxia-inducible factor signaling in pheochromocytoma: turning the rudder in the right direction. *Journal of the National Cancer Institute*, 105(17), pp.1270–1283.

K. Boyd, B. Korf, A. Theos, 2009. Neurofibromatosis type 1. *J Am Acad Dermatol.*, 61(1), pp.1–16.

Kaelin, W.G. & Ratcliffe, P.J., 2008. Oxygen sensing by metazoans: the central role of

the HIF hydroxylase pathway. *Molecular cell*, 30(4), pp.393–402.

Kansanen, E. et al., 2013. The Keap1-Nrf2 pathway: Mechanisms of activation and dysregulation in cancer. *Redox Biology*, 1(1), pp.45–49.

Katsanis, N. et al., 2001. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science (New York, N.Y.)*, 293(5538), pp.2256–9.

Khanna, H. et al., 2009. A common allele in RPGRIP1L is a modifier of retinal degeneration in ciliopathies. *Nature genetics*, 41(6), pp.739–745.

Knudson, A.G., 1971. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, 68(4), pp.820–823.

Kondo, K. et al., 2002. Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein. *Cancer cell*, 1(3), pp.237–246.

Korpershoek, E., 2011. The Pathogenesis of Pheochromocytomas: Of Mice and Men. Erasmus University, Rotterdam.

Latif, F. et al., 1993. Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science (New York, N.Y.)*, 260(5112), pp.1317–1320.

Laurier, V. et al., 2006. Pitfalls of homozygosity mapping: an extended consanguineous Bardet-Biedl syndrome family with two mutant genes (BBS2, BBS10), three mutations, but no triallelism. *European journal of human genetics : EJHG*, 14(11), pp.1195–1203.

Lepoutre-Lussey, C. et al., 2016. From Nf1 to Sdhb knockout: Successes and failures in the quest for animal models of pheochromocytoma. *Molecular and cellular endocrinology*, 421, pp.40–48.

Liem, K.F. et al., 2009. Mouse Kif7/Costal2 is a cilia-associated protein that regulates Sonic hedgehog signaling. *Proceedings of the National Academy of Sciences of the*

United States of America, 106(32), pp.13377–13382.

Liu, J. et al., 2015. Targeting the ubiquitin pathway for cancer treatment. *Biochimica et biophysica acta*, 1855(1), pp.50–60.

Lorenzo, F.R. et al., 2013. A novel EPAS1/HIF2A germline mutation in a congenital polycythemia with paraganglioma. *Journal of molecular medicine (Berlin, Germany)*, 91(4), pp.507–512.

Louie, C.M. et al., 2010. AHI1 is required for photoreceptor outer segment development and is a modifier for retinal degeneration in nephronophthisis. *Nature genetics*, 42(2), pp.175–180.

Luchetti, A. et al., 2015. Profiling of somatic mutations in phaeochromocytoma and paraganglioma by targeted next generation sequencing analysis. *International journal of endocrinology*, 2015, p.138573.

Maher, E.R, 2013. HIF2 and endocrine neoplasiaL an evolving story. *Endocrine related cancer*, 20(3):C5-7.

Maher, E.R., 2014. Phaeochromocytoma and paraganglioma: next-generation sequencing and evolving Mendelian syndromes. *Clinical medicine (London, England)*, 14(4), pp.440–444.

Majewski, J. et al., 2011. What can exome sequencing do for you? *Journal of medical genetics*, 48(9), pp.580–589.

Martinez, V.D. et al., 2014. Unique Pattern of Component Gene Disruption in Complex in Serous Ovarian Cancer. , 2014.

Martinez, Victor; D; Vucic, Emily A; Thu, Kelsie L; Pikor, Larissa A; Lam, Stephen; Lam, W.L., 2014. Disruption of KEAP1/CUL3/RBX1 E3-ubiquitin ligase complex components by multiple genetic mechanisms: Association with poor prognosis in head and neck cancer. *Head & Neck*, 55(7), pp.691–696.

Marxsen, J.H. et al., 2004. Hypoxia-inducible factor-1 (HIF-1) promotes its degradation

by induction of HIF-alpha-prolyl-4-hydroxylases. *The Biochemical journal*, 381(Pt 3), pp.761–767.

Maxwell, P.H. et al., 1999. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature*, 399(6733), pp.271–275.

McIntyre, J.C., Williams, C.L. & Martens, J.R., 2013. Smelling the roses and seeing the light: gene therapy for ciliopathies. *Trends in biotechnology*, 31(6), pp.355–363.

Mendel, G. & Bateson, W., 1865. Experiments in Plant Hybridization (EN transl.). *Journal of the Royal Horticultural Society*, 26(1865), pp.3–47.

Michels, E. et al., 2007. Detection of DNA copy number alterations in cancer by array comparative genomic hybridization. *Genetics in medicine : official journal of the American College of Medical Genetics*, 9(9), pp.574–584.

Mitelman, F., Johansson, B. & Mertens, F., 2007. The impact of translocations and gene fusions on cancer causation. *Nature reviews. Cancer*, 7(4), pp.233–245.

Nagy, R., Sweet, K. & Eng, C., 2004. Highly penetrant hereditary cancer syndromes. *Oncogene*, 23(38), pp.6445–6470.

Nakane, T. & Biesecker, L.G., 2005. No evidence for triallelic inheritance of MKKS/BBS loci in Amish Mckusick-Kaufman syndrome. *American journal of medical genetics. Part A*, 138(1), pp.32–34.

National Human Genome Research Institute, National Human Genome Research Institute (NHGRI). Available at: https://www.genome.gov

Niemann, S. & Müller, U., 2000. Mutations in SDHC cause autosomal dominant paraganglioma, type 3. *Nature genetics*, 26(november), pp.268–270.

Oudijk, L. et al., 2014. H-RAS mutations are restricted to sporadic pheochromocytomas lacking specific clinical or pathological features: data from a multi-institutional series. *The Journal of clinical endocrinology and metabolism*, 99(7), pp.E1376–80.

Pasche, B. & Yi, N., 2010. Candidate gene association studies: successes and failures. *Current opinion in genetics {&} development*, 20(3), pp.257–261.

Pemov, A. et al., 2014. Genetic modifiers of neurofibromatosis type 1-associated caf{é}-au-lait macule count identified using multi-platform analysis. *PLoS genetics*, 10(10), p.e1004575.

Pickart, C.M., 1998. Targeting of Substrates to the 26s proteasome. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, pp.1055–1066.

Pollard, P.J. et al., 2005. Accumulation of Krebs cycle intermediates and over-expression of HIF1α in tumours which result from germline FH and SDH mutations. *Human Molecular Genetics*, 14(15), pp.2231–2239.

Putoux, a. et al., 2012. Novel KIF7 mutations extend the phenotypic spectrum of acrocallosal syndrome. *Journal of Medical Genetics*, 49, pp.713–720.

Putoux, A. et al., 2011. KIF7 mutations cause fetal hydrolethalus and acrocallosal syndromes. *Nature genetics*, 43(6), pp.601–606.

Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D., 2011. RAS oncogenes: weaving a tumorigenic web. *Nature reviews. Cancer*, 11(11), pp.761–774.

Qin, Y. et al., 2010. Germline mutations in TMEM127 confer susceptibility to pheochromocytoma. *Nature genetics*, 42(3), pp.229–233.

Racker, E., 1974. History of the Pasteur effect and its pathobiology. *Molecular and cellular biochemistry*, 5(1-2), pp.17–23.

Ratcliffe, P.J., 2013. Oxygen sensing and hypoxia signalling pathways in animals: the implications of physiology for cancer. *The Journal of physiology*, 591(Pt 8), pp.2027–42.

van Reeuwijk, J., Arts, H.H. & Roepman, R., 2011. Scrutinizing ciliopathies by unraveling ciliary interaction networks. *Human molecular genetics*, 20(R2), pp.R149–57.

Rocha, S., 2007. Gene regulation under low oxygen: holding your breath for transcription. *Trends in biochemical sciences*, 32(8), pp.389–397.

Rous, P., 1911. A Sarcoma of the Fowl Transmissable by an Agent Separable from the Tumour Cells. *Journal Experimental Medicine*, pp.397–411.

Rowbotham, D.A. et al., 2014. Multiple Components of the VHL Tumor Suppressor Complex Are Frequently Affected by DNA Copy Number Loss in Pheochromocytoma. *International journal of endocrinology*, 2014, p.546347.

Rykalina, V.N. et al., 2014. Exome Sequencing from Nanogram Amounts of Starting DNA: Comparing Three Approaches T. Gilbert, ed. *PLoS ONE*, 9(7), p.e101154.

Sadowski, M. et al., 2012. Protein monoubiquitination and polyubiquitination generate structural diversity to control distinct biological processes. *IUBMB life*, 64(2), pp.136–142.

Scheffzek, K. et al., 1997. The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science (New York, N.Y.)*, 277(5324), pp.333–338.

Schinzel, A., 1979. Postaxial polydactyly, hallux duplication, absence of corpus callosum, macrencephaly and severe mental retardation: a new syndrome? *Helvetica paediatrica acta*, 34, pp.141–146.

Selak, M. a. et al., 2005. Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-α prolyl hydroxylase. *Cancer Cell*, 7(January), pp.77–85.

Sementchenko, V.I. & Watson, D.K., 2000. Ets target genes : past , present and future.

Shawky, R.M., 2014. Reduced penetrance in human inherited disease. *Egyptian Journal of Medical Human Genetics*, 15(2), pp.103–111.

Sheerin, U., 2014. The the Use of Next of Generation Parkinson ' s Sequencing Technologies to Dissect A etiologies disease and Dystonia. , (March), pp.1–225.

Shimobayashi, M. & Hall, M.N., 2014. Making new contacts: the mTOR network in metabolism and signalling crosstalk. *Nature reviews. Molecular cell biology*, 15(3), pp.155–62.

Sims, D. et al., 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), pp.121–32.

Singla, V. & Reiter, J.F., 2006. The primary cilium as the cell's antenna: signaling at a sensory organelle. *Science (New York, N.Y.)*, 313(August), pp.629–633.

Siva, N., 2015. UK gears up to decode 100 000 genomes from NHS patients. *The Lancet*, 385(9963), pp.103–104.

Sporn, M.B. & Liby, K.T., 2012. NRF2 and cancer: the good, the bad and the importance of context. *Nature Reviews Cancer*, 12(8), pp.564–571.

Stratton, M.R., Campbell, P.J. & Futreal, P.A., 2009. The cancer genome. *Nature*, 458(7239), pp.719–724.

Teeuw, M.E. et al., 2014. Consanguineous marriage and reproductive risk: attitudes and understanding of ethnic groups practising consanguinity in Western society. *European journal of human genetics : EJHG*, 22(4), pp.452–457.

Toledo, R.A. et al., 2013. In vivo and in vitro oncogenic effects of HIF2A mutations in pheochromocytomas and paragangliomas. *Endocrine-related cancer*, 20(3), pp.349–359.

Tucker, T., Marra, M. & Friedman, J.M., 2009. Massively parallel sequencing: the next big thing in genetic medicine. *American journal of human genetics*, 85(2), pp.142–154.

Varela, I. et al., 2011. Exome sequencing identifies frequent mutation of the SWI/SNF

Complex Gene PBRM1 in renal carcinoma. *Journal of Urology*, 186(3), p.1150.

Verykokakis, M. et al., 2007. The RAS-dependent ERF control of cell proliferation and differentiation is mediated by c-Myc repression. *Journal of Biological Chemistry*, (282, 30285-30294).

Vogelstein, B. et al., 2013. Cancer Genome Lanscapes. *Science*, 339(6127), pp.1546–1558.

Wake, N., 2013. *Identification and Functional Analysis of a Novel Renal Cell Carcinoma (RCC) Susceptibility Gene from an RCC associated Constitutional Chromosomal Translocation . By Naomi Catherine Wake A thesis submitted to the University of Birmingham for the degree*.

Wake, N.C. et al., 2013. UBE2QL1 is disrupted by a constitutional translocation associated with renal tumor predisposition and is a novel candidate renal tumor suppressor gene. *Human mutation*, 34(12), pp.1650–1661.

Wallace, M.R. et al., 1990. Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science (New York, N.Y.)*, 249(4965), pp.181–186.

Walsh, T. et al., 2010. Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *American Journal of Human Genetics*, 87(1), pp.90–94.

Wang, G.L. et al., 1995. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proceedings of the National Academy of Sciences of the United States of America*, 92(12), pp.5510–5514.

Warburg, O., 1956. Injuring of Respiration the Origin of Cancer Cells. *Science*, 123(3191), pp.309–14.

Weizmann Institute of Science, The GeneCards Human Gene Database. *1996*, p.Tripartite Motif Containing 21.

Wessel, J. et al., 2015. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nature communications*, 6, p.5897.

Xie, C. et al., 2014. Ubiquitin-conjugating enzyme E2C: a potential cancer biomarker. *The international journal of biochemistry {&} cell biology*, 47, pp.113–117.

Ye, Y. & Rape, M., 2009. Building ubiquitin chains: E2 enzymes at work. *Nature reviews. Molecular cell biology*, 10(11), pp.755–764.

Zaghloul, N.A. et al., 2010. Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 107(23), pp.10602–7.

Zhang, D.D., 2010. The Nrf2-Keap1-ARE signaling pathway: The regulation and dual function of Nrf2 in cancer. *Antioxidants {&} redox signaling*, 13(11), pp.1623–1626.

Zhang, J. et al., 2011. The impact of next-generation sequencing on genomics. *Journal of genetics and genomics = Yi chuan xue bao*, 38(3), pp.95–109.

Zheng, B. & Fisher, D.E., 2015. Metabolic vulnerability in melanoma: a ME2 (me too) story. *The Journal of investigative dermatology*, 135(3), pp.657–659.

Zhu, M. & Zhao, S., 2007. Candidate gene identification approach: progress and challenges. *International journal of biological sciences*, 3(7), pp.420–427.

Zhuang, Z. et al., 2012. Gain-of-Function Mutations in Paraganglioma with Polycythemia. *New England Journal of Medicine*, 367(10), pp.922–930.

Zlotogora, J., Penetrance and expressivity in the molecular age. *Genetics in medicine : official journal of the American College of Medical Genetics*, 5(5), pp.347–352.

Gene	Primer Sequence	Product Size	Annealing Temp
KIF7	F: CTCTGCCTGCCCTCCTG	323bp	58.2°C
	R: GTGTGCTGGACACTCCCC		56.0°C
RALGAPA2	F: GAGCAGGACAAAAGCCATTC	282bp	54.2°C
	R: CCAATTCTATTTGTTGGCGTGT		56.4°C
RYK	F: GCTTCTGCTTGCGCTGTT	229bp	50.1°C
	R: CACGGCTCGCAGACTGAT		50.1°C
TMEM158	F: CTCCAATGCTTCAGTCAACG	212bp	54.2°C
	R: AGAGCAGTAGGTCGCACTGG		55.8°C
ARHGEF5	F: TGAATGCCCCACACTTAACA	373bp	53.4°C
	R: GGGTCCTCATCTTCTTGAGC		57.5°C
CA12	F:CAGGTCCTGGGGTTAGATGA	344bp	57.4°C
	R: AGGATGATGGGGACTGAGGT		57.4°C
CAPS2	F:GAGGGCTATAAGCAGTATAGTCAGAG	376bp	60.82°C
	R: CCCAAGGTACGTGAAAAACTTC		56.0°C
CNTN6	F: ACAGGTGTGATGGGGGAATA	381bp	55.4°C
	R: GTGCTGCCTAAATGGGTCTC		57.5°C
ERF	F: GGATGAGGGTGAGGACAAGA	374bp	57.5°C
	R: CCACTTTGCCCAGAGCTTAG		57.5°C
PCGF5	F: AAGGTGTGGCAACCAAGTTC	368bp	55.4°C
	R: CCCTCTTTCTACTACCATGTAGCTC		60.1°C
UBE2C	F: GGCTGAGGCAAGAGAATCAC	381bp	57.5°C
	R: GCTCAACCGAGGCTTAATTT		53.4°C
HIF2A Ex.9	Primers from Graeme Clark		
HIF2A Ex. 12	F:TGGAATAGTGTTTGTGAGGTCGT	678bp	43.0°C
	R: GGGCTAAATGGGGTATCAGA		50.0°C
HRAS EXON 2	F:ATTTGGGTGCGTGGTTGA	420bp	66.2 °C
	R:CCTCTAGAGGAAGCAGGAGACA		64.3 °C
HRAS EXON 3	F:AGAGGCTGGCTGTGTGAAC	450bp	69.1°C
	R:GCAGCGGCATCCAGGACAT		68.3°C
KRAS EXON 2	F:GTGTGACATGTTCTAATATAGTCA	450bp	57.4°C
	R:GAATGGTCCTGCACCAGTAA		63.2°C
KRAS EXON 3	F:CGTCATCTTTGGAGCAGGAA	500bp	57.9°C
	R:ACTCCACTGCTCTAATCCCC		58.5°C
NRAS EXON 2	F:GCTAAGGATGGGGGTTGCTA	374bp	61.0°C
	R:TCCGACAAGTGAGAGACAGG		58.9°C
NRAS EXON 3	F:GGACAAACCAGATAGGCAGAA	384bp	59.2°C
	R:CCTCATTTCCCCATAAAGATTC		58.8°C

Supplementary Table 1. Table of primers and their conditions not included in main body of text.

Gene	Exon	Polymorphism	dbSNP Ref	MAF
UBE2C	1	p.Ser23Arg	rs11537645	G=0.0433/217
ERF	4.i	p.Arg205Arg	rs11557114	C=0.0855/428
ERF	4.ii	p.Arg205Arg	rs11557114	C=0.0855/428
ERF	4.iii	p.Ala410Ala	rs35493131	G=0.1016/509
ERF	4.iii	p.Ser327Ser	rs61735151	A=0.0779/390
ME2	13	p.Gly450Glu	rs649224	A=0.1562/782
KEAP1	4	p.Leu471Leu	rs1048290	G=0.4918/2463
KEAP1	4	p.Asp479Asp	rs35074907	A=0.0096/48
CUL2	13	p.Thr386Thr	rs16935840	T=0.1494/748
CUL2	20	p.Ser650Ser	rs12830	T=0.2837/1421

Supplementary table 2. Information From dbSNP on the Reported Polymorphisms Identified in this Study. *The information includes the polymorphisms identified, dbSNP reference and the minor allele frequencies for each variant.*

Gene	Exon	Sporadics Sequenced	Familials Sequenced
CUL2	2	31	29
CUL2	3	55	83
CUL2	4	20	64
CUL2	5	40	23
CUL2	6	40	23
CUL2	7	21	32
CUL2	8	55	72
CUL2	9	55	72
CUL2	10	55	83
CUL2	11	55	83
CUL2	12	20	28
CUL2	13	55	83
CUL2	14	32	70
CUL2	15	32	70
CUL2	16	22	83
CUL2	17	55	83
CUL2	18	31	81
CUL2	19		
CUL2	20	34	74
CUL2	21	25	83
CUL2	22	55	83
CUL3	1	22	35
CUL3	2	48	26
CUL3	3	40	83
CUL3	4	31	47
CUL3	5	27	37
CUL3	6	55	77
CUL3	7	20	22
CUL3	8	50	71
CUL3	9	47	72
CUL3	10	42	29
CUL3	11		
CUL3	12	48	61
CUL3	13		
CUL3	14	52	34
CUL3	15	48	83
CUL3	16	27	54

Supplementary table 3. A table summarising the numbers of patients sequenced for each exon of CUL2 and CUL3. *Dark green boxes indicate that no sequencing was obtained for these exons.*