Genotype-Phenotype Variability in Birt-Hogg-Dubé syndrome

By Mark Devaney

15th September 2013

A thesis submitted to the University of Birmingham
for the degree of MASTERS OF PHILOSOPHY

CLINICAL AND EXPERIMENTAL MEDICINE

COLLEGE OF MEDICAL AND DENTAL SCIENCES

UNIVERSITY OF BIRMINGHAM

SEPTEMBER 2013

Abstract:

Birt-Hogg-Dubé Syndrome (**OMIM:** 135150) is the result of heterozygous germline mutations within the Folliculin gene on 17p11.2 and presents with a classic clinical triad of symptoms: Kidney neoplasia, spontaneous pneumothorax and fibrofolliculomas (benign skin growths). BHD displays no correlation between the type of mutation of the Folliculin gene and the phenotype; patients with the same mutations can present with disparate phenotypes. There is a known age-dependant penetrance present in BHD. Patients with BHD syndrome have an increased risk of Renal Cell Carcinoma (RCC) and associations with Colorectal cancers have been found. This study aims to investigate the potential of this genotype-phenotype variability in two ways. Firstly, it uses Sanger sequencing to test patients for germline polymorphisms known to be associated with susceptibility to renal tumours or melanomas to detect a correlation between possible "genetic modifiers" and risk of skin and renal tumours in BHD syndrome. Secondly investigating the genetic changes involved in renal tumourigenesis in BHD syndrome and sporadic kidney cancer. None of the high risk candidate genes were associated with increased risk of tumours in BHD patients. Sequencing of a BHD sporadic kidney tumour implicated ARHGEF5 in both BHD and sporadic renal tumourigenesis.

Contents

Table of Figures

List of Tables:

Introduction

Cancer (formally known as a malignant neoplasm) is a highly variable spectrum of diseases with over 200 distinct forms, that can occur across almost any tissue type within the body (UK, 2011). Cancer is widely characterised as uncontrolled proliferation of cells via abnormal mitosis, this rapid growth can form malignant tumours that compete with regular tissues for resources to the detriment of the organism as a whole. To initiate transformation of a regular cell into a cancer cell – Carcinogenesis, a mutation must occur that affects cell division, differentiation and/or regulation. This event effectively enhances the cell's ability to divide and grow, providing a clonal advantage to all cells that inherit that mutation allowing somatic evolution of those cells by natural selection. As these cells grow and divide they out-compete appropriately regulated cells surrounding them, and continue to acquire additional secondary mutations that further enhance their ability to grow. Under normal conditions cell division is tightly regulated, tumour-suppressor genes function by preventing carcinogenesis, as a class of genes they can inhibit cell division, initiate apoptosis or other protective properties of a cell that might promote their survival. For carcinogenesis to proceed, a tumour-suppressor gene is typically lost, or a mutation results in a reduced expression of the gene. In contrast to the loss of a tumour-suppressor, a proto-oncogene can become mutated into an oncogene which enables cells to avoid apoptosis and continue to replicate. Activation of a proto-oncogene occurs primarily by three mechanisms. Direct damage to the DNA in the gene itself or the surrounding regulatory region, chromosomal abnormalities - translocations or duplications can both result in increased expression or dysregulation. Lastly differences in mRNA expression and stability have a profound effect on gene expression, resulting in increased expression of oncogene proteins (Ross, 1995). A "two-hit" hypothesis was proposed by Knudson; tumour-suppressor genes are typically

recessive, if one allele is lost the second copy of the allele might be able to compensate though with reduced expression. Therefore a “second hit” is normally required to remove the second alleles’ expression and allow carcinogenesis to progress (Knudson, 1971).

The ability of cancerous cells to invade other tissues and regions of the body is known as metastasis, whilst previously characteristic of late-state malignant tissues recent research suggests this ability is not unique to cancer (Podsypanina, et al., 2008). Metastasis is an invasive process that typically leads to life-threatening complications and secondary or "metastatic" tumours affecting different organ systems.

Transformation of a regular cell into a cancer cell has distinct hallmarks as codified by Hanahan & Weinberg. They described six essential processes that effectuate transformation into a tumourigenic cell and the eventual malignancies that develop (Hanahan & Weinberg, 2000). Mitogenic growth signals and factors for that cell must become self-sufficient with lesser dependence on external stimulants, driving further cell growth. Cancer cells can produce their own growth factors independently to sustain their proliferation or they can induce "regular" cells in the surrounding micro-environment to secrete growth factors via reciprocative signalling (Cheng, et al., 2008). Additionally the cancer cells must acquire a resistance or insensitivity to inhibitory anti-proliferative signals from surrounding cells to prevent growth disrupting homeostatic mechanisms within the tissue. Under normal circumstances apoptotic mechanisms drive programmed cell death via extracellular and intracellular signals to remove defective or deleterious cells via cell-suicide usually in response to stress or damaging conditions. The *TP53* gene encoding p53 is the renowned "guardian of the genome" and a well characterised tumour-suppressor gene that can counteract carcinogenic processes and instability. It accomplishes this by inducing

apoptosis, activating DNA repair mechanisms, or arresting the cell cycle growth which allows for repair mechanisms to act before cell growth is allowed to continue (Bunz, et al., 1998). It can also inhibit angiogenesis - a hallmark discussed below (Dameron, et al., 1994). Naturally *TP53* is lost in most cancers either by deletion, inactivation or upstream/downstream signalling in mutant cells and is a significant factor in carcinogenesis (Bourdon, 2007). Cells undergo senescence (biological ageing) and cease dividing after approximately fifty cell cycles - this apparent limit to cell division is known as the "Hayflick limit" (Hayflick & Moorhead, 1961) whilst they no longer divide they can remain active in a metabolic sense thereby contributing to phenotype. Senescent cells have been implicated in certain diseases and age-related disorders and can disrupt nearby cells and cellular functions (Baker, et al., 2011). Cells can also undergo senescence in response to environmental conditions or DNA damage, the shortening of telomeres at the end of a chromatid is also considered DNA damage by the cell. Telomere shortening is a natural process that prevents truncation during cell division by allowing the telomere itself to shorten at equivalent rates between different cell types (Daniali, et al., 2013). Evidence in mice suggests that when deliberately engineered to lack Telomerase (the enzyme responsible for elongation of telomeres) premalignant cells likely to develop into specific tumour types undergo senescence and subsequently apoptosis (Artandi & DePinho, 2010). For a proto-cancer cell to continue to mutate it must evade these apoptotic and senescent signals and processes. Telomerase is typically upregulated in malignant cells, maintaining and extending the telomeres - effectively gaining biological immortality and limitless replicative potential. Left unchecked these cells will divide continuously until they are limited by the availability of oxygen and nutrients constraining the size of the tumour; for the transformation to continue cells must acquire angiogenesis to form new blood vessels from existing ones allowing further

propagation (McDougall, et al., 2006). In contrast to developmental processes during embryogenesis, in adults angiogenesis and vasculogenesis are relatively dormant processes. Vasculogenesis is the formation of endothelial progenitor cells that form the interior lining of blood vessels (the endothelium), these cells are also used during angiogenesis and are heavily implicated in metastasis and the continued growth of tumours (Gao, et al., 2008). Angiogenesis in adults is transiently activated for regeneration and wound healing and the female menstrual cycle, under normal conditions angiogenic processes are turned "off" and tightly regulated. During tumourigenesis the "angiogenic switch" must become activated permanently; the continued formation of blood vessels can supply the increasing metabolic requirements of the growing neoplasia (Hanahan & Folkman, 1996). Compounding these physiological changes, the precursor cells must acquire invasive capabilities to spread to other neighbouring tissues. Modification and loss of cell-cell adhesion molecules (CAMS) and cadherins that normally allow cells to bind together and maintain tissue structure contribute to the invasive abilities of malignant cells (Tepass, et al., 2000). Upregulation of extracellular protease expression and downregulation of their respective inhibitors allow invasion across epithelial cell barriers and across blood vessels. These proteases are also involved in both angiogenic processes and mitogenic growth signalling which further contribute to expansion and invasion (Stetler-Stevenson, 1999). Late-stage metastatic tumours display heterogeneity within their cell population; individual cells possess differing capabilities for metastasis including no potential at all in a small subset (Kozlowski, et al., 1984). This heterogeneity results in inefficiency within the metastatic process - only a small number of cells are viable (Weiss, 1990). The heterogeneity of the growing tumour is a non-linear independent process that produces great diversity of potentially metastatic cells within the tumour micro-environment; this also results in a reduced effectiveness of

biomarker prognoses and therapeutic measures (Kuukasjarvi, et al., 1997). A study in renal cell carcinoma shows that metastases can differ almost completely from the primary tumour due to multiple severe genomic alterations this indicates that detection of metastases based upon the primary tumour's mutations alone is limited (Bissig, et al., 1999).

These six hallmark features both encourage and are a result of genomic instability from the loss of tumour-suppressor genes and other "caretaker" systems that prevent, detect or repair DNA damage, as well as activation of proto-oncogenes, lesser fidelity during replication or environmental factors providing the correct set of conditions for carcinogenesis. The non-deterministic nature of random mutation means the pre-requisite mutations and physiological changes can occur at different points and time frames within the process not necessarily in the described order - an alternate pathway to cancer but ultimately these hallmarks are common in most types of cancer. There are exceptions to this theory; in Leukaemia, the leukocytes already possess invasive capabilities as part of their function and thus malignant leukocytes do not need to acquire them.

Recent evidence published by the Cancer Genome Atlas Research network suggests tumour progression in clear cell Kidney cancer is further driven by a change in tumour metabolism - “Metabolic shift", shifting from one metabolic pathway towards another. These changes in metabolism are also heralded by significant epigenetic changes, both in sporadic and non-sporadic cases and include mutations within HIF (Hypoxia-Inducible factors) factors such as *HIF1A* and *EPAS1*. Chromatin modification as a result of frequent mutations in chromatin modifier genes in ccRCC (Clear cell renal cell carcinoma) alongside histone methylation pathways result in significant changes throughout the tumour furthering its progression. The severity and aggressiveness of a ccRCC tumour can be correlated with these shifts in the

metabolism of the tumour and the prognosis of the patient (The Cancer Genome Atlas Research Network, 2013).

Birt-Hogg-Dubé Syndrome (**OMIM:** 135150) is characterised by heterozygous germline mutations within the *Folliculin (FLCN)* gene on 17p11.2 and the disease presents with a classic clinical triad of symptoms: Kidney neoplasia, spontaneous pneumothorax and fibrofolliculomas (benign skin growths). Approximately 15-30% of patients with BHD present with renal tumours (Linehan, et al., 2003); predominantly the histology is that of a chromophobe tumour (23% of BHD tumours) or clear cell subtypes, hybrid tumours of chromophobe and oncocytoma (epithelial tumours derived from oncocytes noted for a granular cytoplasm) are also highly prevalent (Palmirotta, et al., 2010). Chromophobe tumours do not stain well and are noted for their pale histology under microscopy, these tumours often form in the collecting ducts of the kidney. Chromophobe tumours are distinct from "clear-cell" RCC (which account for approximately 70% of sporadic renal tumours) which derive from the proximal tubules and stain distinctly (Zhou & He, 2013). BHD syndrome displays an age-dependant penetrance, the majority of people with kidney cancer are over the age of 55. BHD typically displays no correlation between the type of mutation in the Folliculin gene and the phenotype; patients with the same mutations can present with disparate phenotypes. However, evidence suggests there may be a phenotype-genotype correlation between BHD and colorectal polyps and colorectal cancer (Khoo, et al., 2002) (Nahorski, et al., 2010), in this instance both mutations are on exon 11 at c.1285. Evidence suggests allelic heterogeneity of BHD can be a factor in some colorectal neoplasias. The syndrome has also been associated with melanoma risk via activation of the mTOR pathway, which when unregulated promotes cell growth and melanoma formation. (Toro, et al.,

1999) (Cocciolone, et al., 2010). There is mounting evidence suggesting a possible association with BHD syndrome and an increase risk of Breast cancer syndromes (Palmirotta, et al., 2010).

Fibrofolliculomas (benign skin tumours of the hair follicle usually found on the face), trichodiscomas (a related type of benign facial skin tumour) and acrochordons (known as 'skin tags' are harmless flesh-coloured growths that protrude out of the skin in the neck, groin, or armpits) were originally described as the characterising triad of skin manifestations that partially define BHD syndrome. Later definitions describe acrochordons as common within a population, instead focusing on fibrofolliculomas (FFs) and have been expanded to occasionally include angiofibromas (which are more characteristic of tuberous sclerosis), no early childhood onset for FFs has been reported for confirmed *FLCN* mutations (Menko, et al., 2009). Fibrofolliculomas have traditionally been classified as benign hair-follicle tumours (hence the name) however, recent evidence suggests this classification is incorrect and that they are more accurately considered tumours derived from sebaceous glands. It has been argued that they represent abnormal differentiation and proliferation from the “sebaceous mantle” and progenitor sebaceous gland stem cells driven by abnormal hormonal imbalances. This explanation may also aptly explain the noted frequency of FFs in sebaceous gland rich areas (For example the nose) over other parts of the skin (Vernooij, et al., 2013). Fibrofolliculomas are the most common BHD associated symptom and are found in over 80% of patients. There have been several studies that suggest an association between BHD syndrome and an increased risk of skin tumours - both melanoma and basal cell carcinomas however, it has been difficult to demonstrate a direct link between these them (Toro, et al., 1999) (Khoo, et al., 2002) (Menko, et al., 2009) (Mota-Burgos, et al., 2013).

Melanomas are derived from melanocytes –photoreceptors within the skin that secrete melanin in response to DNA damage and exposure to UV-radiation which shield neighbouring cells from damaging UV rays. Melanocytes are unusual in that they rely on external growth factors/mitogens and do not produce their own therefore their proliferation is dependent on cell signalling and mitogenic factors produced by other cells. Mitogen-Activated Protein Kinases/Extracellular Signal-Regulated Kinase (MAPK-ERK) pathways are integral in melanoma progression and are driven by ras oncoproteins. The MAPK-ERK pathway activates multiple growth enhancing factors, and is inhibited by *Gng12;* expression of *Gng12* is notably decreased in cases of primary or metastatic melanoma allowing sustained cell growth by preventing ras inhibition in contrast to upregulated MAPK-ERK pathways (Byrum, et al., 2013). Malignant melanomas are infrequent skin cancers but are aggressive and often metastasize, they are noted for their lethality (Jerant, et al., 2000).

Basal cell carcinomas (BCC) are among the most common skin cancers; BCCs develop from progenitor cells within the basal-skin layer and displays a greater frequency within fair-skinned or pale individuals. In contrast to a melanoma, a BCC rarely metastasizes and is much less dangerous in comparison to a melanoma. Evidence suggests basal cell carcinomas use a less efficient energy pathway than a melanoma and are actively hindered by the immune system preventing aggressive invasion and metastasis commonly found in melanoma; neither neoplasms are under hypoxic conditions. Within the oxidative phosphorylation pathway, the key enzyme in ATP generation responsible for driving this pathway is found up-regulated in melanoma but not basal cell carcinomas allowing a greater amount of cell activity and aggressively metastatic pancreatic cancers. A lower less efficient metabolism and interference from the immune system would explain the much less harmful

and invasive basal cell carcinoma despite displaying shared energy pathways and mutations with melanomas (Xu, et al., 2012).

Folliculin *(FLCN)* is identified as the gene responsible for Birt-Hogg-Dubé syndrome (Nickerson, et al., 2002) and is located on chromosome 17p11.2, however its function is currently unknown. The Folliculin protein is implicated in energy-sensing and signalling pathways such as mTOR (via negative regulation), AMPK and RhoA (Nahorski, et al., 2012) (Reiman, et al., 2012). Evidence suggests it acts as a tumour-suppressor gene, albeit atypically -whilst bi-allelic inactivation is required for the formation of renal tumours, the development of skin nodules and fibrofolliculomas only appear to require inactivation of one allele whilst retaining a functional copy resulting in haploinsufficiency (in this context a single allele cannot produce enough functional protein by itself to result in a wildtype phenotype) (van Steensel, et al., 2007). Recent evidence suggests folliculin protein contributes to VHL-mediated suppression of tumour growth similarly to *VHL*'s tumour suppression and autophagy (cell self-degradation during low energy conditions as a survival mechanism) regulation activity. This suggests FLCN can contribute towards VHL tumour-suppression by affecting autophagic processes without mTOR involvement although this mechanism is not well understood (Bastola, et al., 2013). Additionally the *FLCN* homolog (*dFLCN)* in Drosophila when down-regulated resulted in upregulation of rRNA synthesis in contrast to overexpression of *dFLCN* which had the opposite effect, this supports the idea of *FLCN* as a tumour-suppressor showing that it can negatively regulate rRNA (ribosomal-RNA) synthesis (Gaur, et al., 2013). There are over 132 known mutations in the FLCN gene; FLCN appears to have a hyper-mutatable tract on exon 11, this tract is responsible for 50% of BHD mutations in affected families. The majority of these variants are truncating mutations

caused by insertions/deletions thus causing a frameshift mutation. This results in the premature termination of the FLCN protein which in the absence of nonsense-mediated decay, can result in BHD syndrome (Lim et al 2009). These mutations in the FLCN gene are detected in 88% of patients with BHD (Toro, et al., 2008); it is possible that the remaining 12% may be the result of unidentified modifier effects or inadequacies of current diagnostic methods and sequencing. Mutations of Folliculin in sporadic chromophobe renal tumours and oncocytic tumours are notably infrequent; this would suggest that tumourigenesis has additional pre-requisites in addition to loss of FLCN. Additional mutations and/or inactivating mutations are a possible aetiology (Nagy, et al., 2004); furthermore FLCN shows a wide expression profile across various tissues. FLCN mRNA shows high expression in most tissues yet displays reduced expression in cell types typically associated with the disease phenotype, namely skin, lung and kidney tissues. These findings are consistent with the idea that FLCN functions as a tumour-suppressor in renal cancers (Warren, et al., 2004).

Whilst BHD is known to display age-dependant penetrance; the lack of a phenotype-genotype correlation may suggest that modifier alleles and epistatic mechanisms play a significant role in the development of the BHD phenotype. There is a precedent for this in Von Hippel-Lindau syndrome (**OMIM:** 193300, henceforth known as VHL), another familial cancer syndrome with an autosomal dominant inheritance. *VHL* is a tumour suppressor gene which predisposes to multiple neoplasms most notably Renal Cell Carcinomas (RCC). Mutations of *VHL* have been reported to be in up to 75% in sporadic Clear Cell RCC (ccRCC). *VHL* is also responsible for other malignant and benign tumours affecting other organ systems such as pancreatic tumours and cysts, phaeochromocytomas and haemangioblastomas (Foster, et al., 1994) (Gnarra, et al., 1994). Contrary to Birt-Hogg-Dubé

syndrome, VHL displays phenotype-genotype correlations (Pavlovich & Schmidt, 2004). Epigenetic silencing can be a more prevalent cause compared to somatic mutations in some tumour suppressor genes. Methylation and other epigenetic repression are implicated as a factor in sporadic VHL, hereditary non-polyposis colorectal cancer, breast and ovarian cancer (Banno, et al., 2012). Hypermethylation of the CpG islands in the VHL tumour-suppressor gene was found to result in inactivation of VHL in Renal Cell Carcinoma tumour subsets consistent with this model (Herman, et al., 1994). CpG islands are cytosine and guanine rich sites common in mammalian promoter regions found in up to 40% of genes (Fatemi, et al., 2005), they are methylated age-dependently (Banno, et al., 2012). Such methylation is capable of acting as the "second-hit" for tumourigenesis progression in inherited cancer syndromes as defined using Knudson's Two-hit model. Inactivation of the second allele of a vital tumour-suppressor gene, a "Loss of Heterozygosity" is a major event in initiating and driving tumourigenesis. Atypical promoter methylation results in epigenetic silencing and a loss of expression. In the context of a tumour-suppressor gene this provides a clonal advantage, mirroring the selective advantage beneficial mutations might provide a cell or organism. Naturally this results in a positive feedback effect, predisposing to additional changes within that cell population both genetic and epigenetic. (Baylin & Ohm, 2006). Clonal Expansion of this nature has been found in multiple cancer types, including the Kidney (Bardessy, et al., 1995). There is evidence of this occurring in von Hippel-Lindau syndrome (Prowse, et al., 1997); it is possible that this may also occur in Birt-Hogg-Dubé Syndrome. There is significant evidence to suggest that passenger deletions can occur to surrounding genes during a homozygous deletion of a tumour-suppressor gene, resulting in ancillary functional losses. It has been estimated that up to 11% of coding genes responsible

for the production of proteins are lost in this manner in human cancers (Muller, et al., 2012).

Loss of heterozygosity and Methylation of the promoter regions in Folliculin are more prevalent causes of inactivation in sporadic BHD cases in contrast to somatic mutation (Khoo, et al., 2003). In an investigation of the involvement of Folliculin in primary sporadic RCC and colorectal cancers, several pathogenic somatic *FLCN* mutations were found. Inactivation of BHD in sporadic clear-cell RCC only occurred in a subset of samples. In addition an absence of epigenetic silencing of *FLCN* was detected in multiple patients alongside a lack of hypermethylation of promoter regions in primary RCC patients. (Fernandes da Silva, et al., 2003). *FLCN* promoter methylation and SNPs associated with modifying risk of a certain phenotype (Single Nucleotide Polymorphisms are variations that don't have a major effect on gene function in contrast to a mutation) are demonstrated to have an effect on tumour suppression alongside TP53 in sporadic chromophobe Renal Cell Carcinomas (Gad, et al., 2007). A recent study has further associated VHL with BHD syndrome, demonstrating a direct connection between *FLCN* tumour-suppression in VHL and autophagy regulation. This is further supported by the ability of *VHL* to suppress HIF factor activity, necessary for autophagic processes. In clear-cell RCC cells with *VHL* loss, lower levels of *FLCN* proteins can be detected even without mutations present within *FLCN* suggesting further interaction between *VHL* and *FLCN.* Clear-cell RCC can be considered a sub-type of BHD-related cancer so these findings are of particular interest; it has been suggested microRNAs produced by VHL might affect or repress translation of *FLCN* proteins (Bastola, et al., 2013).

This study focuses on the potential modifier effects that might influence and predispose to specific phenotypes within BHD syndrome and the genetic events implicated in renal tumourigenesis in BHD and sporadic renal tumours. Germline genetic variations Single Nucleotide Polymorphisms that can affect cell structure, growth or regulation can predispose an individual to specific cancers and can also modify the risk of cancer in patients with inherited cancer syndromes. Thus germline Single Nucleotide Polymorphism (SNP) variants of the BRAF gene were identified as predisposing males risk by up to 4% to malignant melanomas. This increase is greater than the risk conferred by the known melanoma susceptibility locus for *CDKN2A* (codon 599 - V5993 on exon 15) increases risk of melanoma cases by less than 1%. *CDKN2A* mutations are detected and associated with approximately 25% of familial cases of melanoma. Germline *BRAF* SNPs are more commonly found in the population than somatic *CDKN2A* mutations. (Meyer, et al., 2003). Whilst the individual risk per variant might not be significant on its own in combination with other variants their cumulative increase of neoplasms can be significant. SNPs tend to result in a relatively small increase in risk but are far more common within a given population in comparison to an inherited susceptibility mutation. Recently there have been studies investigating other modifiers in Renal Cell Carcinoma; in particular a study by Vahteristo used a genome-wide linkage analysis to compare loci between RCC and Hereditary Leiomyomatosis (HLRCC) (Vahteristo, et al., 2010). No evidence for a genetic modifier was found in those loci however the study did not refute the existence of potential modifiers in RCC. Evidence suggests the SNPs within *CCDN1* (Cyclin D1) modifies risk in germline VHL tumour expression alongside HIF-independent mechanisms (Zatyka, et al., 2002). *CCND1* and its hypoxia-inducible factor (HIFs) have also been associated with a susceptibility locus at *11q13.3* in renal cancer which is covered in detail in a later section (Schödel, et al., 2012).

Germline SNPs in several genes (including *MMP1, MMP3* and *VEGF*) have been demonstrated to act as modifiers for RCC risk for sporadic and familial VHL disease (Ricketts, et al., 2009).

Modifier effects have also been studied in other familial cancer syndromes; variants within the promoter region of *BRCA1* contribute to risk of familial breast and ovarian cancer risk (Bielinksa, et al., 2013). Another study found a modifier locus at 6p24 specific to women carrying *BRCA2* mutations that modified breast cancer risk but had no effect on patients with *BRCA1* mutations or the general population (Gaudet, et al., 2013). Several low penetrance alleles have been studied and linked to familial breast cancer, no link was found between these SNPs and sporadic breast cancer cases (Cardeñosa, et al., 2012).

Aims:

The aim of this study was to further investigate the marked Phenotype-Genotype variability displayed in Birt-Hogg-Dubé syndrome, in particular to investigate candidate modifier alleles for potential effects on skin and tumour risks in BHD syndrome. In addition, tumourigenesis is a multistep process and although inactivation of *FLCN* is the first step in renal tumourigenesis in BHD syndrome, other genetic events occur. To date the nature of the alterations and how they compare to those in sporadic RCC has not been identified. The two elements of my MPhil studies were:

1) To analyse polymorphic variation at candidate modifier genes in 70 BHD patients DNA to see if there is any correlation between the genotypes at these loci and the patients' phenotypes.
2) To undertake bioinformatic analysis on exome data from two BHD renal cancers and then analyse additional sporadic RCC tumour DNA samples to determine whether

genes implicated in BHD renal tumourigenesis might also be implicated in sporadic RCC. If common somatic mutations occurred in specific genes in both BHD and sporadic RCC then those genes could be considered as candidate modifier genes for risk of developing kidney cancer in BHD patients.

Selection of Polymorphic variants:

To select which alleles might be implicated as Modifier Alleles a literature search was performed through recent and relevant studies. The selection criteria for relevant polymorphisms included: frequency of allele in the population, the nature of association with a relevant characteristic of BHD (melanomas, neoplasia risks) the calculated relative risk and P-value of the polymorphism in the study. The odds ratio measures the association of an outcome such as increased risk of a specific disease versus the 'exposure' which in this context refers to medical history or specific health factors. An odds ratio of greater than 1 is associated with a higher risk of the outcome (a disease phenotype) and conversely an odds ratio less than 1 results in a lower likelihood (Szumilas, 2010). The following eight polymorphisms were selected to analyse.

ASIP (rs4911414[T] and rs1015362[G])

Variants rs4911414[T]and rs1015362[G] are alleles of the *ASIP* gene (Agouti-signalling protein) which is associated with the production and regulation of melanin pigments and is known to bind to *MC1R* (Kanetsky, et al., 2002). Both rs4911414 and rs1015362 are known to consistently modulate melanoma risk in tandem and show an association with a fair skin colour and increased melanoma risk (both present with an OR of 1.45, P-value of 1.2×10^{-9}) even after adjusting for *MC1R* variants (Nan, et al., 2009). *ASIP* is known to affect specific pigmentation traits in humans such as the skin, eye and hair pigmentations. It is also

associated with a multitude of pleiotropic manifestations such as diabetes and obesity. Evidence suggests *ASIP* does not contribute to pigmentation via the same mechanisms as *MC1R* (Voisey, et al., 2002) but that polymorphisms in non-coding regions might result in loss of function. *ASIP* is an antagonist ligand that results in a less protective yellow form of melanin - phaeomelanin (rather than the darker eumelanin) by inhibiting *MC1R* receptors (Voisey, et al., 2003).

ASIP displays a strong association with increased risk of cutaneous melanomas and basal cell carcinomas (BCC) with a combined P-value of 1.2×10^{-9} (Gudbjartsson, et al., 2008).

MC1R rs1805007 [T]

MC1R (Melanocortin-1-receptor gene) is a highly polymorphic gene that encodes melanocyte-stimulating hormone (MSH) which regulates melanin pigment production (Melanogenesis). The gene is a risk factor for melanomas and plays a significant role in pigment variation in humans. Red hair, fair skin and ultraviolet radiation sensitivity are strongly associated with the *MC1R* gene (Perez Oliva, et al., 2009). Melanocortin-1-receptor mutations are common in European populations and are responsible for an increase of Melanoma risk. It is a known modifier of risk in *CDKN21* mutants, which are lower frequency but higher penetrance mutations that confer a greater risk. In combination with *CDKN2A* the penetrance of mutations increased significantly by 84% with a mean onset age of 37.8 years (Box, et al., 2001). *ASIP* variants modify the risk of melanoma and basal cell carcinoma when interacting with *MC1R* (Gudbjartsson, et al., 2008). The rs1805007 variant (T allele) was found to have the highest Basal Cell Carcinoma risk with an OR of 1.55 and a P value of 5.9×10^{-9} ; it is known to be a red hair colour variant (Nan, et al., 2011). The

rs1805007 and rs1805008 variants are common non-synonymous mutations in European populations and have a discernible effect on pigmentation (Makova & Norton, 2005).

Variants of *MC1R* are also known to significantly increase the risk of melanoma in a Latvian population (Ozola, et al., 2013) and the phenotype displays a "gene dosage" effect - carriers of multiple variants display a greater vulnerability. Evidence suggests there is a secondary mechanism for modification of melanoma progression and risk induced by *MC1R* variants that is unrelated to pigmentation (Robinson & Healy, 2002) or it may negate the protection provided by darker pigmentation. This mechanism would help account for discrepancies noted in previous research showing increased melanoma risk attributed to specific variants even when melanoma-associated phenotypes such as fair skin and red hair are not present or otherwise accounted for (Valverde, et al., 1996) (Palmer, et al., 2000). Both *MC1R* variants have been suggested to have been driven by recent positive selection; this might suggest these alleles confer an adaptive beneficial trait (Sulem, et al., 2007). One hypothesis for the benefits involved the fairer skin in European populations allows for increased production of vitamin D in the skin reducing the risk of vitamin D deficiency and ricketts, however there is not enough statistical evidence to support this claim (Rana, et al., 1999). Alternatively it has been argued *MC1R* is neutrally selected and the high polymorphic rate in European populations is due to the "relaxation of constraints" outside of Africa as these variants are likely more deleterious in African climates (Harding, et al., 2000), this does not discount the possibility of diversifying selection however.

TYR rs1126809 (R402Q)

Tyrosinase (TYR) located on 11q14-21 is the gene responsible for the conversion of the non-essential amino acid tyrosine into Melanin via the enzyme tyrosinase. *TYR* is

transcriptionally regulated by *Microphthalmia-associated transcription factor (MITF)*; a family with a heterozygous single base pair deletion in *MITF* were also found to bear the R402Q polymorphism and presented with features consistent with occular albinism and Waardenburg syndrome (Morell, et al., 1997). Polymorphisms within this gene can be attributed to common melanoma risk factors such as freckling and UV sensitivity (Bishop, et al., 2009). (Gudbjartsson, et al., 2008) Found that the *TYR* rs1126809 (A) variant is known to affect eye colour and tanning increased the risk of both Cutaneous Melanoma (OR 1.21, P = $2.8x10^{-7}$) and Basal Cell Carcinoma (OR 1.14, P=$6.1x10^{-4}$). This variant is sometimes referred to as R402Q/1205 G> A this nomenclature describes the amino acid change from arginine (R) to a glutamine (A) at position 402. Eye colour, specifically blue, is known to be a risk factor for melanomas independent of skin pigmentation and nevi count (Bliss, et al., 1995). The A allele of rs1393350 has a strong association with the rs1126809 SNP and blue eye colour (Sulem, et al., 2007) (Allwood & Harbinson, 2013). The rs1393350 variant has also been implicated in melanoma risk with a significance of $2.41x10^{-14}$ (Bishop, et al., 2009). R402Q is also found to significantly increase the risk for squamous cell carcinoma (SCC) as well as melanoma risk (Nan, et al., 2009). Despite the involvement with pigmentation mechanisms, the *TYR* rs1126809 variant bears no association with autosomal recessive ocular albinism (Oetting, et al., 2009) contradicting previous findings (Hutton & Spritz, 2008). However, other mutations in *TYR* are responsible for oculocutaneous albinism type 1 (Ray, et al., 2007). This discrepancy might be resolved with suggested genetic disequilibrium with R402Q - non-random allele association at different loci resulting in a greater population frequency than might be expected through random chance. R402Q substitution is demonstrated to have an effect on the thermo-sensitivity of the protein in albino mice within the ribosome functioning normally at 32 °C transporting melanosomes however at

37°C the protein was shown to misfold and co-localised within the ribosome suggesting a possible mechanism for affecting melanoma risk (Halaban, et al., 2000).

MITF rs149617956 (E318K)

MITF (Microphthalmia-associated transcription factor) is implicated in melanocyte development and differentiation of other cell types. *MITF* regulates *TYR* expression and protein levels by binding to the promoter region. A recent study demonstrated that *MITF* controls pancreatic β cell function via direct regulation of *Pax6* transcription; *Pax6* is also known to be involved in pigment cell differentiation (Mazur, et al., 2013) . The E318K variant was strongly associated with melanoma risk in two independent studies, in both an Australian sample group and a British group. With the British and Australian cases combined, the presence of this *MITF* variant had an OR of 2.19 and a two-sided P value of 0.0003. Notably this variant was also associated with eye colours, specifically non-blue as well as an increased naevus count ("birthmarks"). These features may contribute towards a phenotype that can be associated with the variant most associated with melanoma risk, although linkage and Genome-Wide association studies have had difficulty proving so (Yokoyama, et al., 2011). In the UK sample *Yokoyama et al* noted the extremely low frequency of the "A" allele compared with the G/G genotype, the "A" allele had a frequency of 0.009 in affected patients and a frequency of 0.004 in the control samples. Due to the low frequency of the deleterious allele that it would be harder to implicate it as a causative factor especially in small sample size studies. The phenotypic associations are in contrast to *MC1R* and *TYR* which increase risk in association with "fair skin" traits. Previously *MITF* has been identified as a melanoma oncogene when amplified, leading to greater mortality rates in patients and was more prevalent in metastatic disease (Garraway, et al., 2005).

Furthermore, *MITF* is frequently found mutated or over-expressed in certain melanoma subtypes (Cronin, et al., 2009).

11q.13.3 loci (rs7105934)

A genome-wide association study found an uncommon polymorphism (rs7105934) located on locus 11q13.3 that is implicated in increased renal cell carcinoma (RCC) risk; this locus has no characterised genes. The variant rs7105934 displays an OR of 0.71 and a p-value of $7.8x10^{-14}$ (Purdue, et al., 2011). Risk of renal carcinoma is reduced by 0.69 times when the [A] allele is present, this protective effect seems to be lost or reduced in heavy smokers/drinkers or overweight people. The locus lacks CpG islands which are typically associated with promoter regions, these "Islands" are rich in cytosine and are indirectly methylated as a form of epigenetic regulation (Saxonov, et al., 2006). Methylation of CpG islands in contrast to CpG sites can occur up to 2 kilobases away at highly conserved regions that are termed "CpG island shores" (Irizarry, et al., 2009). In addition to the genome-wide association study, it has also been found that the rs7105934 variant is known to modify RCC risk in a Chinese population (Cao, et al., 2012) with the caveat that additional population studies may be required to validate these findings. The 11q13.3 locus has also been implicated in other human cancers, the rs11228565[A] allele is associated with increased risk of prostate cancer (Gudmundsson, et al., 2009). The locus is commonly amplified in a multitude of cancers including human breast cancers (Lundgren, et al., 2008) and is flanked by genes overexpressed in various cancers such as CCDN1 which is responsible for cell-cycle regulation via encoding cylcin D1 proteins (Turnbull, et al., 2010). rs7105934 is now known to regulate hypoxia inducible factor (HIF) binding using a cyclin D1 transcription enhancer, this can result in a loss of cell-cycle regulation. Evidence suggests this locus contains a long-range HIF-dependant transcriptional enhancer, this enhancer is coupled with the *CCDN1*

promoter which then disrupts HIF binding and hypoxia pathways (Schödel, et al., 2012). *CCDN1* is an oncogene activated by mitogenic deregulation, copy number alterations or infrequently by mutation. The role of Cyclin D1 as a commonly dysregulated oncogenic cell cycle regulator is well known; deactivation of cyclin expression result in the characterised hallmarks of cancer discussed previously and is therefore a potential therapeutic target and biomarker (Musgrove, et al., 2011).

EPAS1 rs11894252

The *Endothelial PAS Domain containing protein 1 (EPAS1)* is located on chromosome 2p21, this gene is known to encode transcription factors that induces genes regulated by low oxygen levels in the cell (Hypoxia) known as Hypoxia-inducible factors (HIF). HIF factors are highly conserved and are essential during development; they can also promote angiogenesis in tumours and other survival factors during hypoxic conditions. HIF factors are commonly upregulated in VHL syndrome as a consequence of mutation to the VHL suppressor proteins (Maina, et al., 2005); overexpression of HIF factors increases the amount of mRNA involved for angiogenesis, energy metabolism and apoptotic pathways during hypoxic conditions as a result *EPAS1* is a key gene for RCC development. *EPAS1* causes increased angiogenesis in RCC by upregulating Vascular endothelial growth factor (*VEGF*) in tumour tissue; both *EPAS1* and *VEGF* mRNAs shown co-expression and are present at significant levels in RCC (Xia, et al., 2001). These HIF factors are a suggested therapeutic target; normal oxygen dependant regulatory processes resume with the reintroduction of pVHL into the tumour (Zimmer, et al., 2004). The *EPAS1* variant rs11894252 located on intron 1 has been shown to increase RCC risk (combined OR 1.14, P value of 1.8×10^{-8}), it was noted the risk was significantly higher in smokers than non-smokers (Purdue, et al., 2011). The apparent differences in risk

between men and women (higher in men) have been mostly attributed to the aforementioned frequency of smoking however; additional confirmation may be required.

MMP1 rs1799750 (G/G)

Matrix Metallopeptidase 1 (Interstitial Collagenase) hence force known as *MMP1* is a gene in the *MMP* gene cluster located on chromosome 11q22.3. The proteins known as matrixins produced by this gene are zinc-dependent proteases that degrade extracellular molecules, as a result *MMP* genes have critical roles in angiogenesis, apoptosis, cell proliferation and differentiation and can drive *in vivo* tumourigenesis (Foley, et al., 2013). A link has been found in *MMP1* mRNA expression between smokers and non-smokers using buttock skin artificially exposed to UV radiation, concluding that tobacco smoking exacerbates ageing effects by inducing *MMP1* expression especially in combination with ultraviolet radiation (Lahmann, et al., 2001). rs1799750 is an insertion polymorphism occurring at position -1607bp located in the *MMP1* promoter, the G/G germline polymorphism is attributed to increased risk in a multitude of cancers. It has been shown to increase risk in ovarian cancer (Kanamori, et al., 1999), colorectal cancer metastasis (Ghilardi, et al., 2001), breast cancer (Hughes, et al., 2007), lung cancer - especially in men and non-smokers (Su, et al., 2006), chronic obstructive pulmonary disease (Wood & Stockley, 2006), chronic kidney disease (Yoshida, et al., 2009) and rheumatoid arthritis onset as a result of polymorphisms in *VEGF* and *MMP1* (Chen & Mattey, 2012). The rs1799750 (2G) variant was determined to have a greater frequency in Renal Cell Carcinoma patients in a Japanese population than population controls might suggest (Hirata, et al., 2004) with an OR of 1.95 CI 1.31-2.91. These findings have been independently verified, with *MMP1* 2G polymorphism being significantly associated with RCC risk confirming the role of *MMP1* and *MMP3* in as susceptibility and risk factors in both familial and sporadic cases. These findings provided an

OR of 1.49 with a 95% CI 1.06-2.08 and a P value of 0.017 (Ricketts, et al., 2009). A recent study determined that by silencing NEU3 (human plasma membrane sialidase enzyme) in RCC its malignancy could be reduced; the silencing resulted in downregulation of *MMP1* and *MMP7* alongside blocking autophagy and inducing apoptosis using the β1 integrin-recycling pathway (Tringali, et al., 2012). This further underscores the role of *MMP1* in RCC development and pathogenesis.

Investigating mechanisms of tumourigenesis in BHD syndrome:

Previously studies in RCC from BHD syndrome have demonstrated somatic *FLCN* mutations consistent with a "two-hit" model of tumourigenesis in patients with germline *FLCN* mutations (Vocke, et al., 2005). However, most examples of human tumourigenesis represent a multistep process and there was no pre-existing information on what additional somatic mutations might be implicated in the development of a renal tumour in BHD patients.

For the second part of the project I had access to exome sequencing data of normal tissue and a kidney tumour from a patient with BHD syndrome who, unusually, had also developed two gastrointestinal tract tumours. I analysed the exome data to identify candidate somatic mutations and then proceeded to validate these in the BHD tumour and then investigated whether relevant genes were mutated in sporadic non-BHD tumours. Exome sequencing data from a second BHD tumour from a Japanese patient with classical BHD syndrome was also investigated to see if there were any common mutations between these two tumours.

Exome sequencing is a more targeted and cost-effective method of sequencing and analysing potential mutations and causative agents for diseases; especially those that display Mendelian inheritance (single locus inheritance in contrast to diseases caused by

multiple gene interactions or loci) by sequencing only the exons of a patient or the "Exome." The significance of the exome is in understanding the functional changes caused by variation within the exons - the coding regions within a gene whereas full genotyping studies also sequence intronic (non-coding) regions (Ng, et al., 2009). The effectiveness of exome sequencing in identifying causes of single gene disorders has already been demonstrated in Miller syndrome (**OMIM:** 263750); a rare autosomal recessive disorder that often results in cleft palate deformations, supernumerary nipples and micrognathia (Ng, et al., 2010). As exons make up 1% of the genome and untranslated regions are often not included, exome sequencing is limited to only finding variation on coding-regions and only variants which that can affect the function of proteins produced by the gene. Non-coding or structural changes that contribute towards a disease phenotype cannot be found through exome sequencing but instead through whole genome sequencing. However, one estimate suggests that exons contain 85% of deleterious causative mutations that significantly affect a disease phenotype (Choi, et al., 2009). Statistical errors such as false positives and negatives are always a problem when generating large amounts of data such as from extensive sequencing via exome/genome techniques however there are methods to reduce these problems.

Exome sequencing is but one tool used in the Cancer Genome Project - an ongoing initiative to identify somatic mutations and variants that are crucial during the acquisition and development of cancer syndromes. The project by the Sanger Institute uses genome-wide analyses and detection to determine critical mutations within tumours using a variety of sequencing methods whether genome, exome or transcriptome (Wellcome Trust Sanger Institute, 2013). This effort has already identified several common mutations and a second

major ccRCC gene (PBRM1) within RCC using exome sequencing (Varela, et al., 2011). The strength of a cancer-genome study has been demonstrated in malignant melanoma showing a predilection for DNA repair processes towards the transcribed regions within a gene over other areas, allowing a wide distribution of acquired mutations over time (Pleasance, et al., 2010).

Methods and Materials

Patient Samples

Patient blood and tumour samples were acquired from patients diagnosed with Birt-Hogg-Dubé syndrome or sporadic RCC (detailed information on age at diagnosis and other clinical characteristics of the patients was not available to me).

Written informed consent had been obtained from all patients and the study had been approved by the South Birmingham Research Ethics Committee. 70 patient samples in total were analysed.. After extraction of DNA by standard methods, the samples were aliquoted in 10µl aliquots and kept as stock solutions in -20°C cold storage. From these solutions, the 'working solutions' were made on a sealed a Star Lab 96 well PCR plate using an assumed concentration of 500ng/µl diluted to 20mg/µl, then made up to 100µl per sample. The plate was sealed using Thermo Scientific Adhesive PCR lids and also stored at -20°C. From this plate 1µl of diluted DNA would be extracted and added to each PCR plate in conjunction with the appropriate primers.

Primer Design

Primers for both stages of the experiment were designed using Primer3 online software with a desired optimum melting temperature of between 58°C and 60°C depending on the sequence. Primer 3 can be found at "www.frodo.wi.mit.edu" alongside documentation of how the software works (Rozen & Skaletsky, 2000). Care was taken to ensure whenever possible that primers had a similar melting temperature (TM) and annealing temperature between pairs and each other as this would allow for greater efficiency by running multiple different primers on the same PCR machine. The GC content (amount of Guanine-Cytosine pairs within a sequence affect the melting temperature as G-C bonds use a triple hydrogen

bond compared to A-T's double bond, as a result more energy is required to break this bond.) was carefully monitored during the design stage. Each oligonucleotide was designed to encompass a region of approximately 500 base pairs (bp) long to provide enough room around the polymorphic site of interest to ensure artefacts created by the sequencing process would not overlap the polymorphism. This was to ensure the sequence in question would be as clear as possible. The oligonucleotides themselves were designed to be 20bp long allowing for enough specificity to bind to the target region. The primer3 software automatically chooses regions within a sequence that are less likely to bind to themselves or to the complementing primer within the mixture forming a distinctive "primer-dimer" when amplified. The oligonucleotides were then ordered from Invitrogen Lifesciences with a synthesis scale of 25nmol, desalted purification and with no additional modifications.

Primer Handling and Storage

The primers were then diluted using purified water, to 200pmol/µl using the given concentrations that came with the primers to create a stock solution. These stock solutions were stored in the -20°C freezer. From these stock solutions a "working solution" could be created using a 1 in 10 dilution; the use of a working solution has multiple advantages. First it reduces the risk of contamination of all subsequent primer solutions by reducing the number of times the "stock solution" is opened and closed. Secondly it allows for further control over the dilution of the primer solution without affecting the stock and thirdly oligonucleotides are susceptible to frequent temperature changes - by freezing and thawing constantly the sequences might become damaged or unreliable. These measures ensure that the stock solution of each primer is kept as sterile and stable as possible, if a working solution becomes contaminated it the problem can be identified easier and can be replaced more easily.

Polymerase Chain Reaction

Each time a set of DNA samples were to be tested for polymorphisms a "master-mix" was created, all components were mixed together into a bulk solution in an eppendorf then aliquot onto a 96-well PCR plate with the sample configuration recorded for later identification. The individual chemicals were kept on ice whilst the mixture was being made. The contents of the master-mix were as follows:

Table 1 PCR reagents

"Master Mix" components, made up to 50µl per Sample.	
Component	**Volume per Sample (µl)**
Buffer (with magnesium)	5
dNTPs	1
Forward Primer	1
Reverse Primer	1
Taq Polymerase	0.2-0.4
Purified Water	40.8

The Taq polymerase was always the last reagent to be added and was kept on ice the entire time. Once the solution had been made the eppendorf containing the master-mix would be vortexed using a Biocote Vortexmixer for a minimum of 3 seconds so the contents would be evenly dispersed within the eppendorf. If this step were not taken it is possible a differing amount of each component would be present in each subsequent aliquot which could affect the quality of the PCR and prevent some wells from amplifying at all. After 49 microlitres of the mastermix were aliquoted into each well using a Gilson pipette, 1µl of the patients' DNA sample would be aliquot from the frozen "working solution" of patient DNA into the

relevant well. Extra care was taken to avoid cross-contamination and ensure throughout the process that any reagent or sample was correctly aliquot into the well, when adding the DNA it was done by eye to observe the sample going into the solution rather than adhering to the side of the wells within the plate. After this the PCR plate would be sealed up with a Thermo Scientific adhesive lid; the seal would be manually checked and firmly wrapped around the PCR plate ensuring it was evenly covering the plate with no room for evaporation to escape the plate whilst being amplified. Once the lid was secure, the PCR plate was inserted into a Biometra Basic T professional PCR machine. The following program was run:

Table 2 PCR protocol

	Polymerase Chain Reaction (PCR) Protocol			
Step	**Temperature (°C)**	**Length (Minutes:Seconds)**	**Go To**	**Loops**
1	95	00:30 (Pre-heating)		
2	95	00:30 (Denaturation)		
3	58-60	00:30 (Annealing)		
4	72	00:45 (Extension)	2	40 cycles.
5	72	07:00 (Extension)		
6	12	Pause (Holding Step)		

The protocol described in Table 2 contains a pre-heating step to 95°C, after this a denaturation-annealing-extension cycle would be replicated forty times before a final extension step lasting 7 minutes. After this the program would transition to a cooling and storage step for when it cannot be manually moved to storage. The temperature in step 3

(the annealing step) is listed as a range; this temperature is changed depending on the melting temperature (TM) of the primers used.

Following this step the plate would either be stored within the PCR machine during the cooling step or transferred to a fridge in preparation for the next step.

Agarose Gel Electrophoresis:

After the plate was amplified within the PCR machine it was important to confirm amplification was a success, therefore testing the samples by running a gel was necessary. 10 µl of each PCR product were transferred using a pipette to a secondary plate for transfer in addition to 5µl of a "loading dye".

The Gel was created using 1.5g of agarose powder dissolved in 150ml of TBE (Tris Acetate EDTA) buffer creating a 1% solution. This solution was heated in a microwave for approximately 2 minutes until the solution became clear to the naked eye. The bottle was then cooled down using tap water around the outside of the bottle until it was comfortable to hold, this takes a minute or more to cool down in this manner. The significance of this step was 1) safety 2) so as not to damage the Gel trays by constant heating and cooling and 3) so it would be cool enough to add in ethidium bromide and reduce the risk of ethidium bromide vapours being released. After the bottle was cooled sufficiently 0.5µl of Ethidium Bromide was added to the solution using a pipette, this functions as a fluorescent tag within the solution once exposed to Ultra-violet light (UV) to visualise the DNA. The bottle was then shaken to make sure the Ethidium bromide was properly mixed, after this it was poured into a gel tray and left to cool for 20-30 minutes, "combs" were inserted into the cooling solution allowing for "wells" to be created for the next step.

Once set the gels were inserted into an electrophoresis tank (Geneflow water bath) submerged in TBE buffer; 15µl would be transferred into the wells created by the combs using a pipette. The contents of each well were documented for later reference and the lid was placed onto the tank. A voltage of 180-200 was run through the gel after checking the correct alignment of the gel so the gel runs forwards not backwards across the current. After 15-20 minutes the gel would be removed from the tank and carefully inserted into a Syngene UV transilluminator. Where it would be exposed to UV light and a photograph of the gel would be taken.

The gels would be checked for bands of amplification with a similar product size predicted by the Primer design; any "Primer-Dimer" formations would be noted as a failed amplification. Whilst the primers were designed to avoid "Primer-Dimer" it was still possible for it to occur. Any samples that had successfully amplified were stored in the fridge alongside the original 96-well PCR plate so if a problem occurred it would be possible to reuse spare PCR product to retry later steps.

"Exosap" clean up phase:

The successful amplicons were recorded and a new 96-well plate was set-up for a cleaning and sequencing preparation stage. Exosap refers to Exon Nuclease I a reagent that improves the quality PCR products by removing the primers not used up in the PCR amplification step whilst the "Shrimp Alkaline Phosphatase" removes any remaining dNTPs (Nucleics, 2013). A new mastermix was created for this stage:

Table 3 “Exosap master mix” Reagents

"Exosap Master Mix" components to 10µl per Sample.	
Exon Nuclease I	0.25µl
Antarctic Phosphatase	1µl
Antarctic Phosphatase Buffer	1µl
Water	1.75µl

The exosap master-mix was vortexed so the contents would be thoroughly mixed.

6µl of the previous successful PCR products were added to the 96-well plate, whilst 4µl of the exosap was carefully pipetted into the well taking care to ensure no product was adhering to the sides of the well so that both the exosap and the product could mix. The plate was tightly sealed with a thermoplastic lid before being placed inside the Biometra. The following "Exosap" protocol was run:

Table 4 Exosap Protocol

Exosap Protocol		
Step	**Temperature (°C)**	**Length (Minutes:Seconds)**
1	37	30:00
2	85	15:00
3	4	Pause

Afterwards the plate was stored in a fridge in preparation for the next step.

The Sequencing Reaction:

For each successful amplicon that had been exosapped; a "forward" and "reverse" well was set up on a fresh 96 well plate and two "master-mixes" were created:

Table 5 Sequencing Reaction Reagents

Sequencing reaction 6µl per Sample.	
"Big dye"	0.75µl
5X Sequencing Buffer	2µl
Forward Or Reverse Primer	1µl
Water	2.25µl

One master-mix would use the Forward primer only and the other would use the Reverse primer only, keeping them separate. In the interests of efficiency multiple successful amplifications of the same polymorphism would be done together so any samples within that set that worked could be pooled together onto one plate. This also attempts to reduce the risk of cross-contamination between different sets or primers, samples and polymorphisms. "Big-Dye" improves the quality of sequencing, but is susceptible to repeated freeze-thaw cycles; therefore it must be aliquoted into smaller working samples. In addition it must be allowed to thaw at room temperature then stored "on-ice" whilst creating the master-mix, as a sequencing dye it is light and UV-sensitive and must be kept wrapped in tinfoil and kept away from windows or lights whilst in use (Applied Biosystems, 2010). These solutions would be vortexed allowing for the reagents to thoroughly mix; then 6µl would be pipette into each well with the forward and reverse solutions kept separate and adjacent. 4µl of the exosap samples would then be pipette into each relevant well both

in the forward and reverse wells. After this the plate would be sealed once more using a thermoplastic lid and run in a PCR machine using the following protocol:

Table 6 Sequencing Reaction Protocol

	Sequencing Reaction Protocol			
Step	**Temperature (°C)**	**Length (Minutes:Seconds)**	**Go To**	**Loops**
1	95	01:00		
2	95	00:30		
3	60	00:30	1	35 cycles.
4	60	04:00		
5	12	10:00		
6	12	Pause (Holding Step)		

The layout of each plate was recorded allowing the well number to be cross-referenced with the sample and primer the well contained. After this protocol had completed the plate would be stored in the fridge once again in preparation for the final "clean-up" step.

Precipitation:

After the plate had undergone the "sequencing reaction" excess reagents and dyes needed to be removed from the solution so there would be greater acuity for the sequenced results.

Using the plate from the previous reaction, 1µl of precipitation was added to each well using a pipette. Following this, 30µl of 100% ethanol was also added to each individual well. The plate was resealed and centrifuged at 4000 rpm (rotations per minute) for 30 minutes. The lid was then removed and the plate was turned upside down onto tissue paper and firmly patted down to remove most of the 100% ethanol, it was then placed back into the

centrifuge and spun upside down with the lid off at 400rpm for 1 minute. The DNA would remain adhered to the bottom of the well in a pellet and would not be dislodged during this process.

200µl of 70% ethanol was then added to each well and once again centrifuged at 4000rpm for 20 minutes; afterwards it would be turned upside down again and patted before being recentrifuged at 400rpm for another minute. These steps were to ensure the supernatant of ethanol and leftover reagent from previous reactions would be completely removed leaving only a pellet of the DNA sample in each well. Without this step the sequencing signal would be almost unreadable, the speed of each cycle was also very important if the speed was not adjusted when the lid was off the DNA pellet would also be lost.

10µl of "Hi-Di formamide" was pipette into each well to resuspend the pellet for sequencing. The plate was then run at 95°C for 5 minutes in a PCR machine to denature the sample. It was then stored on ice until it the plate was sequenced by a lab technician using an Applied Biosystems 3730 DNA analyser.

Analysis:

Analysis of the sequenced samples was performed using a program called "Bioedit (Biological Sequence alignment editor)" available at "http://www.mbio.ncsu.edu/bioedit/bioedit.html" (last checked 4/7/2013). The sequence of the samples would be compared and aligned against the known or expected sequence for that region, differences would be noted including homozygote/heterozygote changes. Statistical analysis was performed using "MedCalc" using a combination of Kaplain-Meier analysis and a logrank test to produce survival curves and data about age of onset on the

censored information. The stated threshold for statistical significance is a P-value of less than 0.05.

Results:

BHD Modifier Study

Analysis of the genotyping data was performed in two steps. Firstly the frequency of homozygotes and heterozygotes at each locus analysed was compared to population data and evidence of Hardy-Weinberg equilibrium was sought. Secondly the effect of individual SNP variants on the major phenotypical features of BHD was determined using Kaplan-Meier analyses.

Population data for each allele was acquired to compare the frequency of the heterozygotes and homozygotes at each locus.

Table 7 Minor-Major Allele frequency population data (Source: 1000genomes.org)

SNP	MAF/MinorAllele Count	Global Frequency		European Frequency	
ASIP rs1015362	T=0.381/829	T = 38%	C = 62%	T = 28%	C = 72%
ASIP rs4911414	T=0.230/502	T = 23%	G = 77%	T = 31%	G= 69%
MITF rs149617956	A=0.001/2	G = 100%	A = 0%	G = 100%	A = 0%
11q.13 rs7105934	A = 0.21	A = 21%	G = 79%	A = 7%	G = 93%
MC1R rs1805007	T = 0.03	T = 3%	C = 97%	T = 8%	C = 92%
TYR rs1126809	A = 0.11	A = 11%	G = 89%	A = 25%	G = 75%
EPAS1 rs11894252	C = 0.44	T = 56%	C = 44%	T = 39%	C = 61%
MMP1 rs1799750	"-" = 0.45	"-" = 45%	C = 55%	"-" = 49%	C = 51%

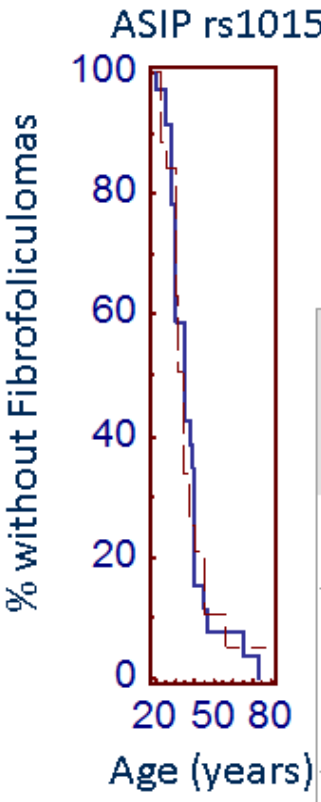
Table 7 uses population data from 1000genomes.org to establish frequency of each variant allele within both a global population and a European population.

Using the program "MedCalc" a Kaplan-Meier logrank analysis was performed on the data sets. The data included in the analysis was: age of onset of fibrofolliculomas and RCC and the type of tumour where applicable from hospital records (though this data was incomplete). This data was then compared using the genotype of the patient and their alleles of interest. The survival and age of onset for each manifestation crossed with genotype produced a series of curves based upon this data; the three types of genotype (two homozygote, one heterozygote) were alternately grouped together to provide a series of curves.

Polymorphisms that had been attributed to increase melanoma risk were analysed against the age of onset for fibrofolliculomas; similarly risk modifiers for RCC were tested against onset of renal tumours. Previously *ASIP* rs4911414[T], rs1015362[G], *MC1R* rs1805007[T], *TYR* rs1126809 [G/A] and *MITF* rs149617956 have been associated with increased risk of developing melanomas.

ASIP rs1015362:

Figure 1: ASIP rs1015362 survival curve for Fibrofolliculoma.



	Factors			
	1		2	
Sample size	35		26	
Median survival	35		34	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	0.971	0.0282	-	-
22	-	-	0.885	0.0627
24	-	-	-	-
25	0.913	0.0482	0.842	0.0725
26	-	-	-	-
28	0.782	0.0732	-	-
30	0.587	0.0883	0.632	0.0980
32	-	-	0.505	0.102
34	-	-	0.421	0.101
35	0.424	0.0889	0.337	0.0967
38	0.385	0.0888	0.253	0.0890
39	0.347	0.0878	-	-
40	0.154	0.0694	0.211	0.0836
45	0.116	0.0618	0.105	0.0672
47	0.0770	0.0519	-	-
56	-	-	0.0527	0.0502
65	0.0385	0.0376	-	-
73	0.000	0.000	-	-
77	-	-	-	-

	Comparison of survival curves (Logrank test)	
Endpoint: Observed n	29.0	22.0
Expected n	29.5	21.5
Chi-square	0.02464	
DF	1	
Significance	P = 0.8753	

Figure 1 displays the survival curve for fibrofolliculomas using a Kaplan-Meier analysis and a logrank statistical test, the C/C variant is considered 1 by the graph with the C/T and T/T variants grouped together as 2. These figures are presented with a combined sample size of 61. Data analysis shows a statistical significance of P= 0.8753 which is above the stated

threshold for statistical significance, therefore there appears to be no difference between alleles.

Figure 2: Survival curve C/C +C/T =1, T/T =2

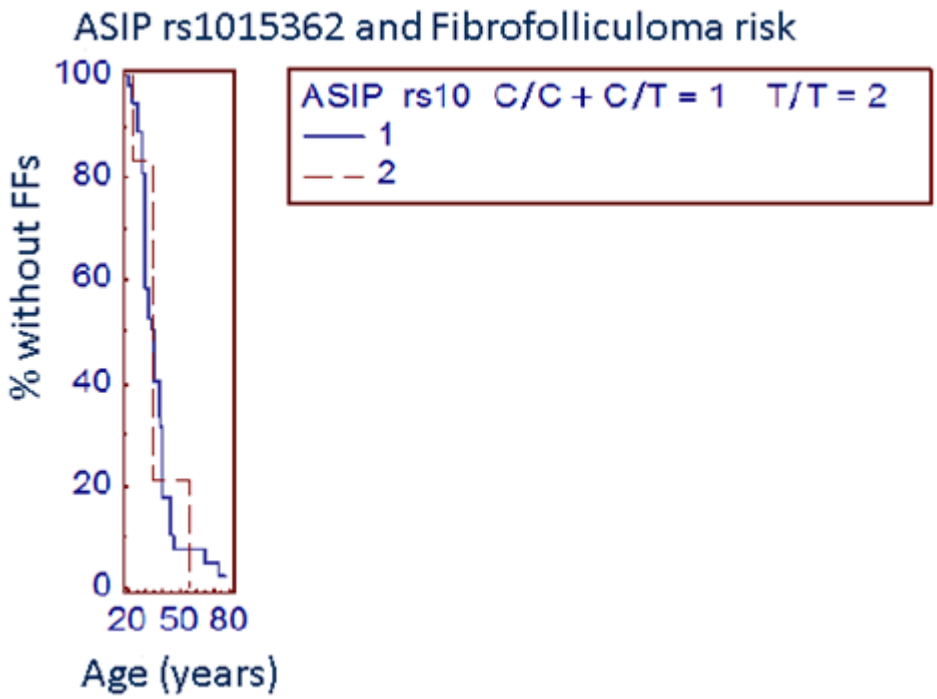


Table 8: Statistical analysis of Figure 2

	Factors			
	1		**2**	
Sample size	55		6	
Median survival	35		35	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	0.982	0.0180	-	-
22	0.945	0.0306	0.833	0.152
24	-	-	-	-
25	0.889	0.0429	-	-
26	-	-	-	-
28	0.808	0.0548	-	-
30	0.586	0.0695	-	-
32	0.525	0.0706	-	-

	Factors			
	1		**2**	
Sample size	55		6	
Median survival	35		35	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
34	0.505	0.0707	0.625	0.213
35	0.404	0.0695	0.208	0.184
38	0.337	0.0679	-	-
39	0.314	0.0670	-	-
40	0.180	0.0565	-	-
45	0.103	0.0466	-	-
47	0.0769	0.0414	-	-
56	-	-	0.000	0.000
65	0.0513	0.0346	-	-
73	0.0256	0.0251	-	-
77	-	-	-	-
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	46.0		5.0	
Expected n	46.2		4.8	
Chi-square	0.01354			
DF	1			
Significance	P = 0.9074			

Hazard ratio	1.0511
95% CI	0.4099 to 2.6955

In figure 2 the same data of 61 patients were analysed with alternate grouping for the variants. Table 8 shows this data has a P-value of 0.9074 and is therefore not statistically significant.

ASIP rs4911414:

Figure 3: Survival Curve for ASIP rs4911414 and age of onset for fibrofolliculomas.

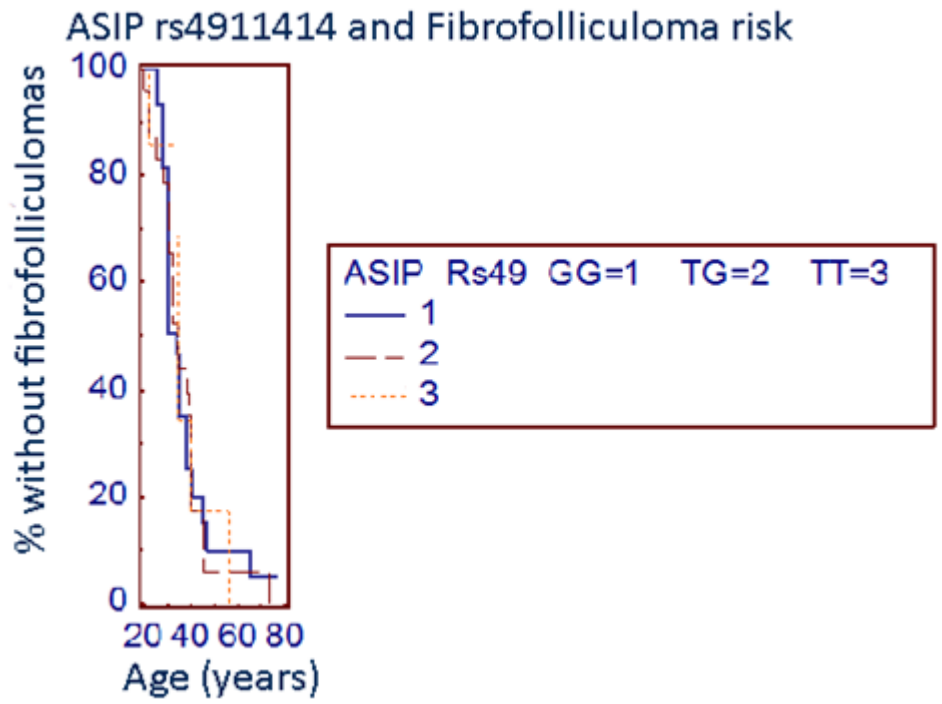


Table 9 Statistical analysis of Figure 3.

	Factors					
	1		2		3	
Sample size	30		24		7	
Median survival	34		35		35	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	-	-	0.958	0.0408	-	-
22	-	-	0.875	0.0675	0.857	0.132
24	-	-	-	-	-	-
25	0.931	0.0471	0.831	0.0770	-	-
26	-	-	-	-	-	-
28	0.815	0.0751	0.787	0.0845	-	-
30	0.504	0.0981	0.656	0.0987	-	-
32	-	-	0.525	0.104	-	-
34	0.466	0.0979	-	-	0.686	0.186
35	0.349	0.0937	0.437	0.103	0.343	0.195

	Factors					
	1		2		3	
Sample size	30		24		7	
Median survival	34		35		35	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error	Survival Proportion	Standard Error
38	0.249	0.0896	0.394	0.102	-	-
39	-	-	0.350	0.0996	-	-
40	0.200	0.0844	0.175	0.0794	0.171	0.156
45	0.150	0.0767	0.0583	0.0545	-	-
47	0.0998	0.0653	-	-	-	-
56	-	-	-	-	0.000	0.000
65	0.0499	0.0481	-	-	-	-
73	-	-	0.000	0.000	-	-
77	-	-	-	-	-	-
	Comparison of survival curves (Logrank test)					
Endpoint: Observed n	23.0		22.0		6.0	
Expected n	23.1		21.7		6.3	
Chi-square	0.01697					
DF	2					
Significance	P = 0.9916					

The P-value quoted in Table 9 is 0.9916 which is above the threshold for statistical significance.

Figure 4 Survival Curve for ASIP rs4911414 and age of onset for fibrofolliculomas

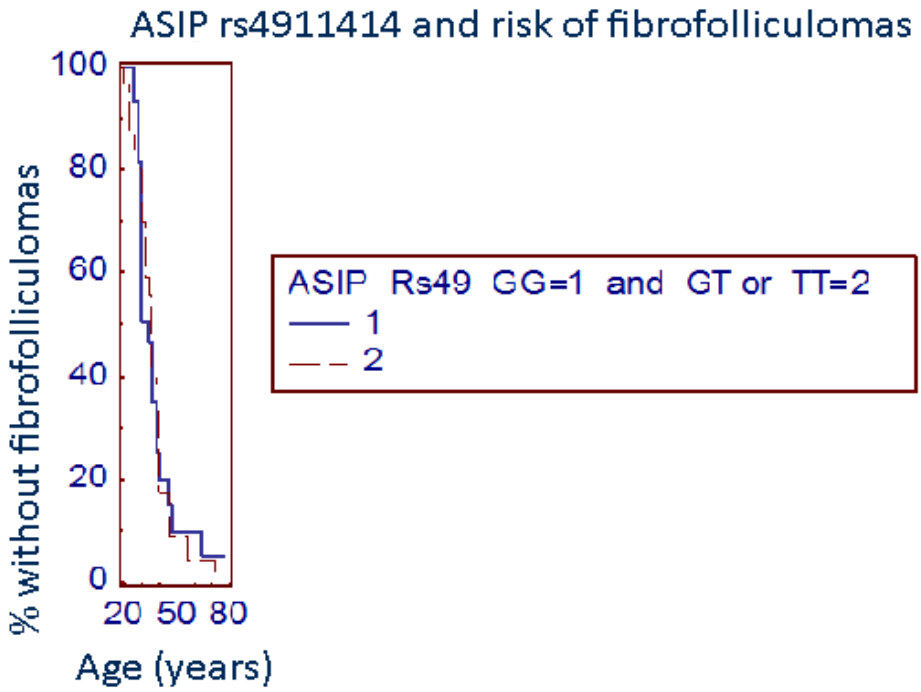


Table 10 Statistical analysis of Figure 4.

	Factors			
	1		2	
Sample size	30		31	
Median survival	34		35	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	-	-	0.968	0.0317
22	-	-	0.871	0.0602
24	-	-	-	-
25	0.931	0.0471	0.836	0.0671
26	-	-	-	-
28	0.815	0.0751	0.801	0.0728
30	0.504	0.0981	0.697	0.0847
32	-	-	0.592	0.0910
34	0.466	0.0979	0.557	0.0921
35	0.349	0.0937	0.418	0.0917
38	0.249	0.0896	0.383	0.0904
39	-	-	0.348	0.0887

	Factors			
	1		2	
Sample size	30		31	
Median survival	34		35	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
40	0.200	0.0844	0.174	0.0707
45	0.150	0.0767	0.0871	0.0561
47	0.0998	0.0653	-	-
56	-	-	0.0435	0.0417
65	0.0499	0.0481	-	-
73	-	-	0.000	0.000
77	-	-	-	-
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	23.0		28.0	
Expected n	23.1		27.9	
Chi-square	0.0007081			
DF	1			
Significance	P = 0.9788			

Hazard ratio	1.0068
95% CI	0.5801 to 1.7475

Table 10 shows a statistically insignificant P-value of 0.9788 suggesting these results lack the statistical power to confirm the hypothesis.

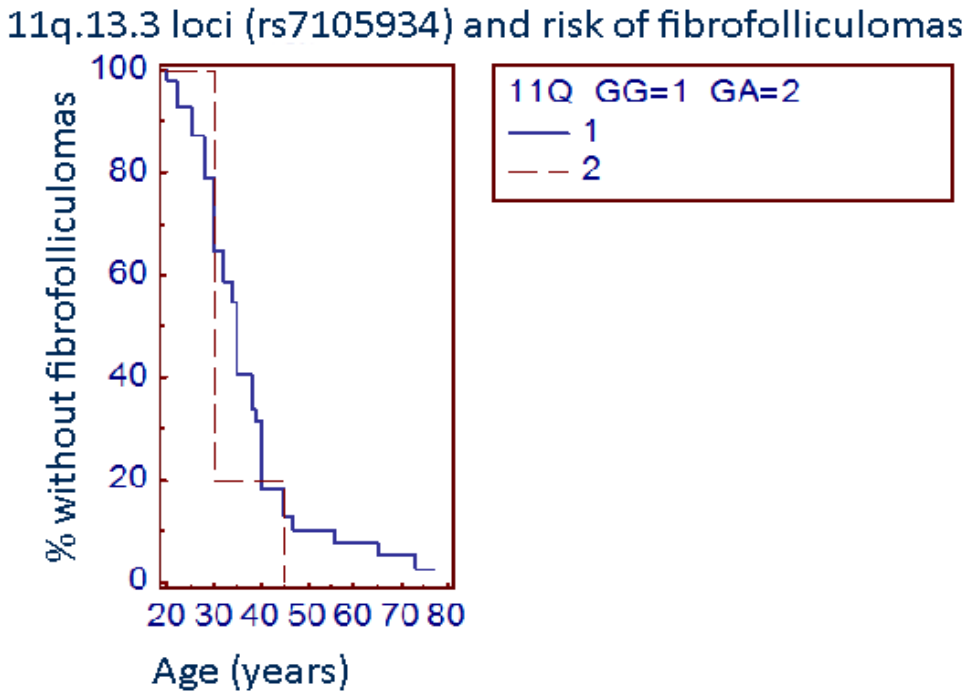
MITF rs149617956 (E318K)

MITF rs149617956 (E318K) was analysed in the 57 patient samples, every sample produced a G/G homozygote with no variation. The A allele has a very low frequency in a given population, (Yokoyama, et al., 2011) reported a frequency of 0.009 in case samples and a

frequency of 0.004 in the control samples so this is to be expected (Chatzinasiou, et al., 2011).

11q.13.3 loci (rs7105934)

Figure 5 Survival Curve for 11q13.3 loci (rs7105934) and age of onset for fibrofolliculomas



There were no A/A genotypes present in the patient data so that cannot be compared for analysis.

Table 11 Statistical analysis of Figure 5

	Factors			
	1		2	
Sample size	56		5	
Median survival	35		30	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	0.982	0.0177	-	-
22	0.929	0.0344	-	-
24	-	-	-	-

	Factors			
	1		2	
Sample size	56		5	
Median survival	35		30	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
25	0.872	0.0453	-	-
26	-	-	-	-
28	0.791	0.0564	-	-
30	0.649	0.0671	0.200	0.179
32	0.588	0.0694	-	-
34	0.547	0.0703	-	-
35	0.405	0.0696	-	-
38	0.338	0.0680	-	-
39	0.315	0.0671	-	-
40	0.180	0.0567	-	-
45	0.129	0.0508	0.000	0.000
47	0.103	0.0467	-	-
56	0.0772	0.0415	-	-
65	0.0515	0.0348	-	-
73	0.0257	0.0252	-	-
77	-	-	-	-
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	46.0		5.0	
Expected n	47.3		3.7	
Chi-square	0.6372			
DF	1			
Significance	P = 0.4247			

The P-value significance quoted in Table 11 is 0.4247 which is above the threshold for statistical significance suggesting these findings are non-significant.

Figure 6 Survival Curve for 11q13.3 loci (rs7105934) and age of onset Renal Cell Carcinoma.

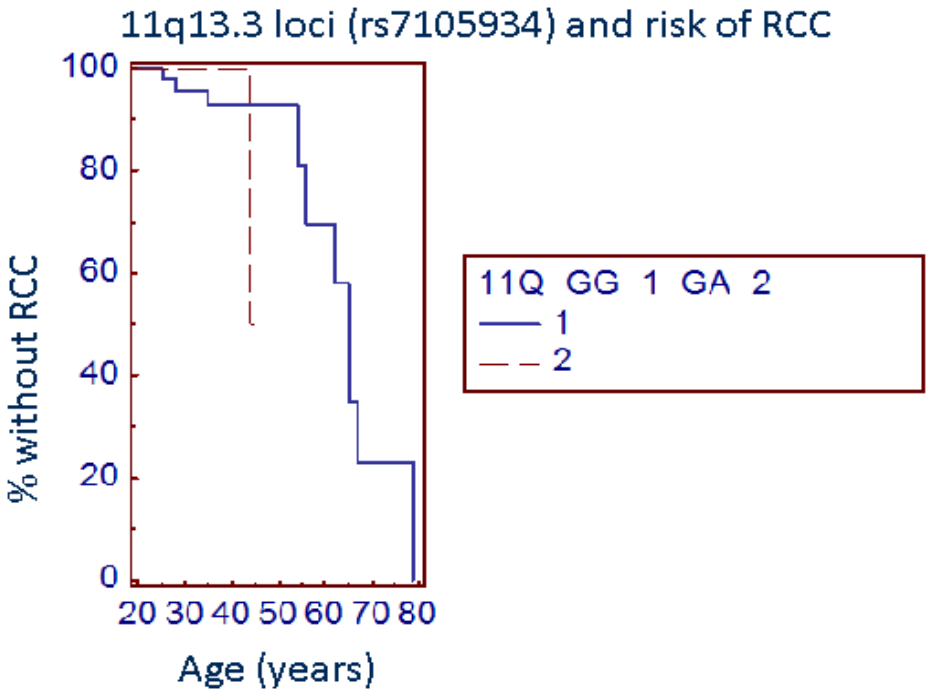


Figure 6: Though there appears to be an earlier age at onset in G/A heterozygotes this did not reach statistical significance and there are insufficient G/A heterozygotes (n=6) to make a reliable interpretation of the results.

Table 12 Statistical analysis of Figure 6

	Factors			
	1		2	
Sample size	58		6	
Median survival	65		-	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	-	-	-	-
22	-	-	-	-
24	-	-	-	-
25	0.981	0.0190	-	-

	Factors			
	1		2	
Sample size	58		6	
Median survival	65		-	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
26	-	-	-	-
28	0.959	0.0285	-	-
30	-	-	-	-
32	-	-	-	-
34	-	-	-	-
35	0.928	0.0411	-	-
38	-	-	-	-
39	-	-	-	-
40	-	-	-	-
44	-	-	0.500	0.354
45	-	-	-	-
47	-	-	-	-
54	0.812	0.114	-	-
56	0.696	0.145	-	-
62	0.580	0.161	-	-
65	0.348	0.160	-	-
67	0.232	0.142	-	-
73	-	-	-	-
79	0.000	0.000	-	-
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	10.0		1.0	
Expected n	10.6		0.4	
Chi-square	0.8808			
DF	1			
Significance	P = 0.3480			

Hazard ratio	2.4909
95% CI	0.1159 to 53.5441

Table 12 suggests statistical non-significance for these findings with a P value of 0.3480 which is above the 0.05 threshold for statistical significance.

TYR rs1126809 (R402Q)

Figure 7 Survival Curve for TYR rs1126809 (R402Q) and age of onset for Fibofolliculomas

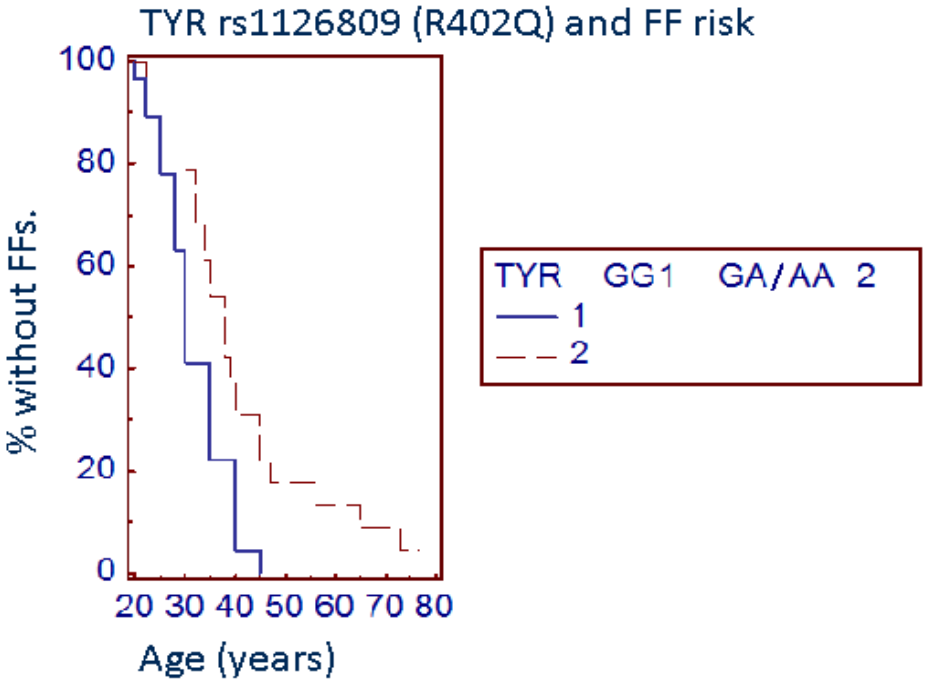


Figure 7 appears to show the G/G homozygote genotype being the most deleterious with an earlier age of onset for Fibrofolliculomas in comparison to G/A and A/A. Previous studies have suggested the A allele modifies risk and increases the risk of both FFs and cutaneous basal cell carcinomas.

Table 13 Statistical analysis of Figure 7

	Factors			
	1		2	
Sample size	28		33	
Median survival	30		38	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	0.964	0.0351	-	-
22	0.893	0.0585	0.970	0.0298
24	-	-	-	-
25	0.781	0.0790	-	-
26	-	-	-	-
28	0.632	0.0926	-	-
30	0.409	0.0947	0.790	0.0765
32	-	-	0.682	0.0878
34	-	-	0.611	0.0921
35	0.223	0.0803	0.539	0.0942
38	-	-	0.423	0.0947
39	-	-	0.385	0.0936
40	0.0446	0.0430	0.308	0.0893
45	0.000	0.000	0.220	0.0827
47	-	-	0.176	0.0769
56	-	-	0.132	0.0691
65	-	-	0.0880	0.0584
73	-	-	0.0440	0.0427
77	-	-	-	-
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	26.0		25.0	
Expected n	17.1		33.9	
Chi-square	8.9989			
DF	1			
Significance	P = 0.0027			

Hazard ratio	0.4851
95% CI	0.2713 to 0.8676

The P-value is 0.0027 which suggests a statistically significant result, that the G/G genotype is the most deleterious contrary to previous findings where the A allele has a greater association in sporadic patients with basal cell carcinoma (Gudbjartsson, et al., 2008).

Figure 8 Survival Curve for TYR rs1126809 (R402Q) and age of onset for Fibofolliculomas

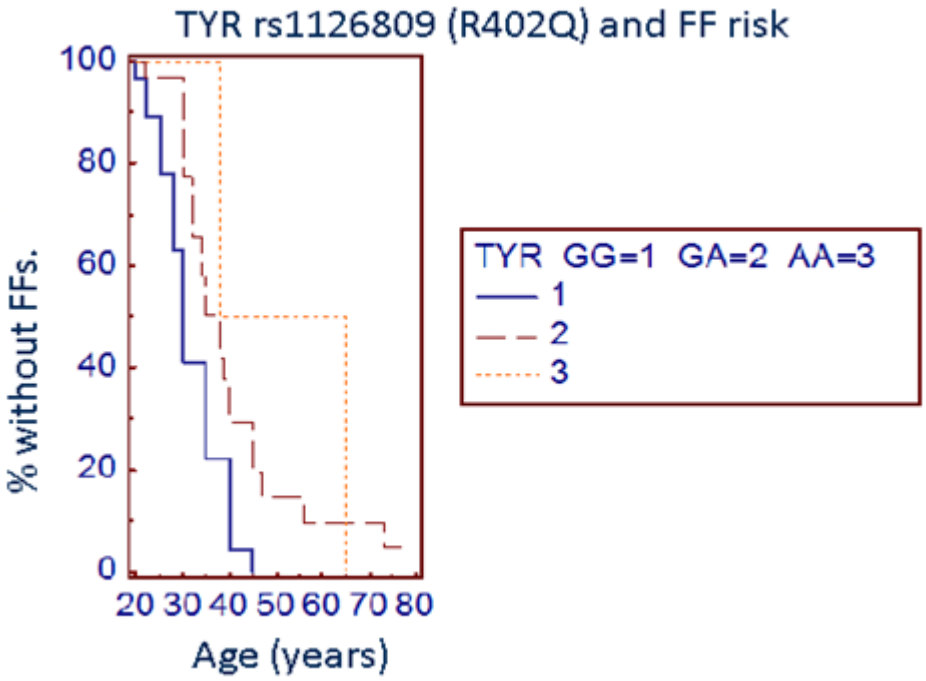


Figure 8 clearly illustrates the G/G genotype being the most deleterious with the A/A genotype (though there were only three patients in this group) being associated with later age of onset for fibrofolliculomas.

Table 14 Statistical analysis of Figure 8.

	Factors					
	1		2		3	
Sample size	28		30		3	
Median survival	30		38		51.5	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	0.964	0.0351	-	-	-	-
22	0.893	0.0585	0.967	0.0328	-	-
24	-	-	-	-	-	-
25	0.781	0.0790	-	-	-	-
26	-	-	-	-	-	-
28	0.632	0.0926	-	-	-	-
30	0.409	0.0947	0.773	0.0817	-	-
32	-	-	0.657	0.0929	-	-
34	-	-	0.580	0.0967	-	-
35	0.223	0.0803	0.503	0.0981	-	-
38	-	-	0.419	0.0980	0.500	0.354
39	-	-	0.377	0.0967	-	-
40	0.0446	0.0430	0.293	0.0916	-	-
45	0.000	0.000	0.195	0.0831	-	-
47	-	-	0.147	0.0754	-	-
56	-	-	0.0977	0.0642	-	-
65	-	-	-	-	0.000	0.000
73	-	-	0.0489	0.0472	-	-
77	-	-	-	-	-	-
	Comparison of survival curves (Logrank test)					
Endpoint: Observed n	26.0		23.0		2.0	
Expected n	17.1		30.2		3.7	
Chi-square	7.1336					
DF	2					
Significance	P = 0.0282					

These contradictory findings also present with a statistically significant P-value of 0.0282.

Figure 9 TYR rs1126809 and RCC risk

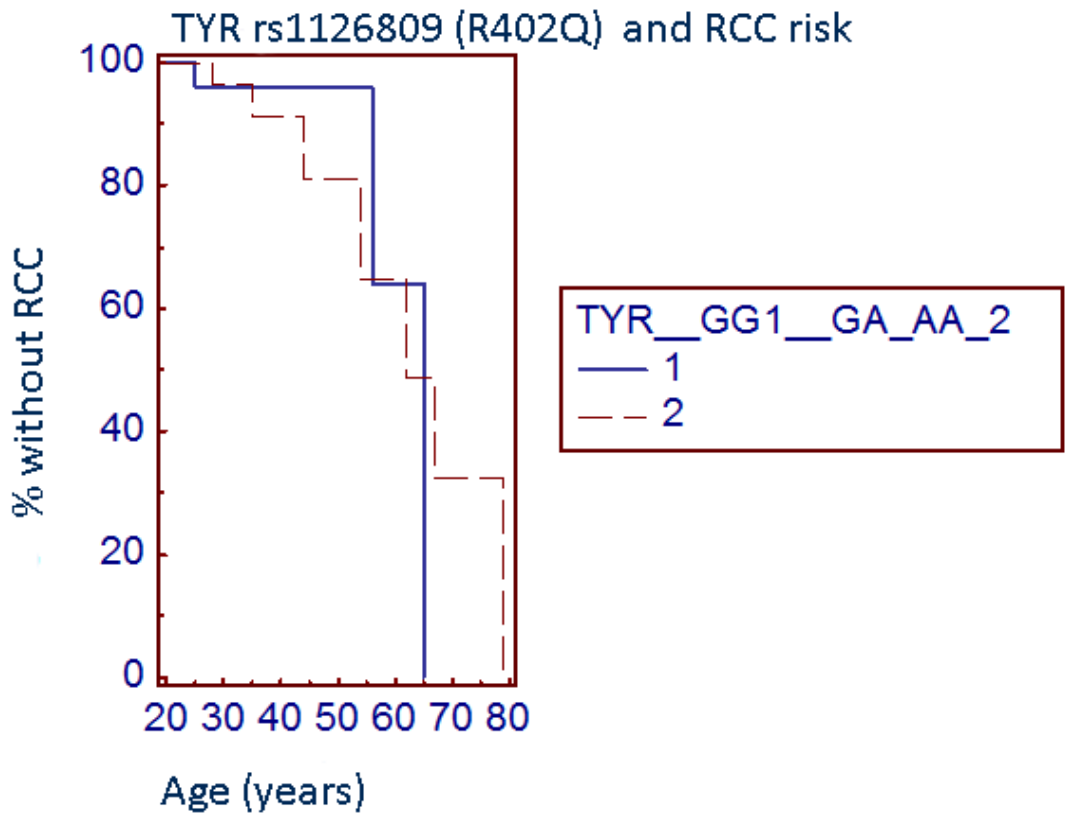


Table 15 Statistical Analysis of Figure 9

	Factors			
	1		2	
Sample size	29		35	
Median survival	65		62	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	-	-	-	-
22	-	-	-	-
24	-	-	-	-
25	0.962	0.0377	-	-
26	-	-	-	-
28	-	-	0.966	0.0339
30	-	-	-	-
32	-	-	-	-
34	-	-	-	-

	Factors			
	1		2	
Sample size	29		35	
Median survival	65		62	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
35	-	-	0.912	0.0612
38	-	-	-	-
39	-	-	-	-
40	-	-	-	-
44	-	-	0.811	0.110
45	-	-	-	-
47	-	-	-	-
54	-	-	0.648	0.170
56	0.641	0.263	-	-
62	-	-	0.486	0.189
65	0.000	0.000	-	-
67	-	-	0.324	0.183
73	-	-	-	-
79	-	-	0.000	0.000
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	4.0		7.0	
Expected n	3.6		7.4	
Chi-square	0.06803			
DF	1			
Significance	P = 0.7942			

There appears to be no statistically significant difference between alleles in terms of RCC risk given the P value quoted in table 10 (P = 0.7942).

EPAS1 rs11894252

Figure 10 Survival Curve for EPAS1 rs11894252 and age of onset for RCC.

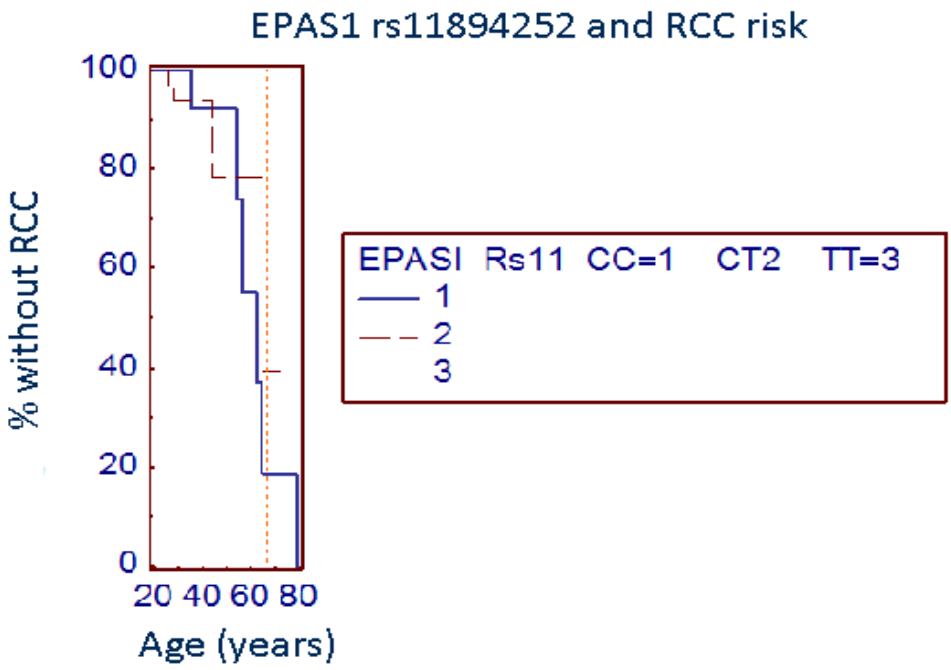


Figure 10 displays the T/T homozygote with the earliest age of onset before age 70, the C/C genotype appears to be the next most deleterious with the C/T genotype having the latest age of onset.

Table 16 Statistical analysis of Figure 9

	Factors					
	1		2		3	
Sample size	21		38		5	
Median survival	62		65		67	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	-	-	-	-	-	-
22	-	-	-	-	-	-
24	-	-	-	-	-	-
25	-	-	0.971	0.0290	-	-
26	-	-	-	-	-	-

	Factors					
	1		2		3	
Sample size	21		38		5	
Median survival	62		65		67	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error	Survival Proportion	Standard Error
28	-	-	0.939	0.0417	-	-
30	-	-	-	-	-	-
32	-	-	-	-	-	-
34	-	-	-	-	-	-
35	0.923	0.0739	-	-	-	-
38	-	-	-	-	-	-
39	-	-	-	-	-	-
40	-	-	-	-	-	-
44	-	-	0.783	0.147	-	-
45	-	-	-	-	-	-
47	-	-	-	-	-	-
54	0.738	0.175	-	-	-	-
56	0.554	0.207	-	-	-	-
62	0.369	0.204	-	-	-	-
65	0.185	0.166	0.391	0.286	-	-
67	-	-	-	-	0.000	0.000
73	-	-	-	-	-	-
79	0.000	0.000	-	-	-	-
	Comparison of survival curves (Logrank test)					
Endpoint: Observed n	6.0		4.0		1.0	
Expected n	5.4		4.1		1.5	
Chi-square	0.2282					
DF	2					
Significance	P = 0.8922					

The P-value from Table 16 for Figure 10 is 0.8922 which suggests statistical non-significance; this would explain the unusual finding of both homozygotes being more deleterious than

the heterozygote. If one allele was noted for having a harmful effect you might expect the heterozygote with the same allele to be more damaging than the homozygote without.

Figure 11 Survival Curve for EPAS1 rs11894252 and age of onset for RCC.

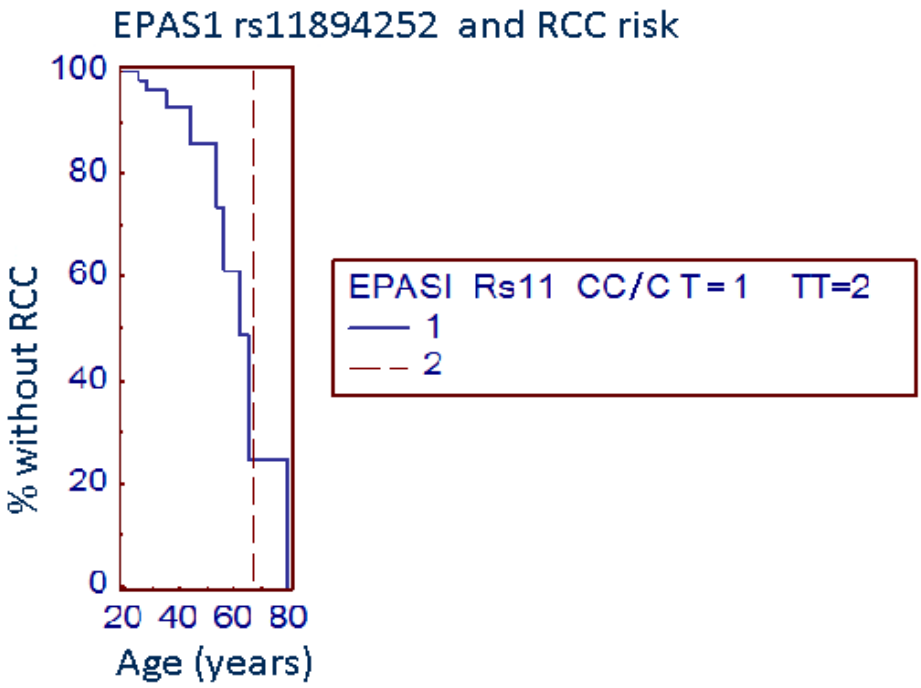


Figure 11 describes a similar trend depicted in Figure 10, with the T/T homozygote being associated with the earliest age of onset. The grouped C/C and C/T genotypes show a later age of onset.

Table 17 Statistical analysis of Figure 11

	Factors			
	1		2	
Sample size	58		6	
Median survival	62		67	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	-	-	-	-
22	-	-	-	-
24	-	-	-	-

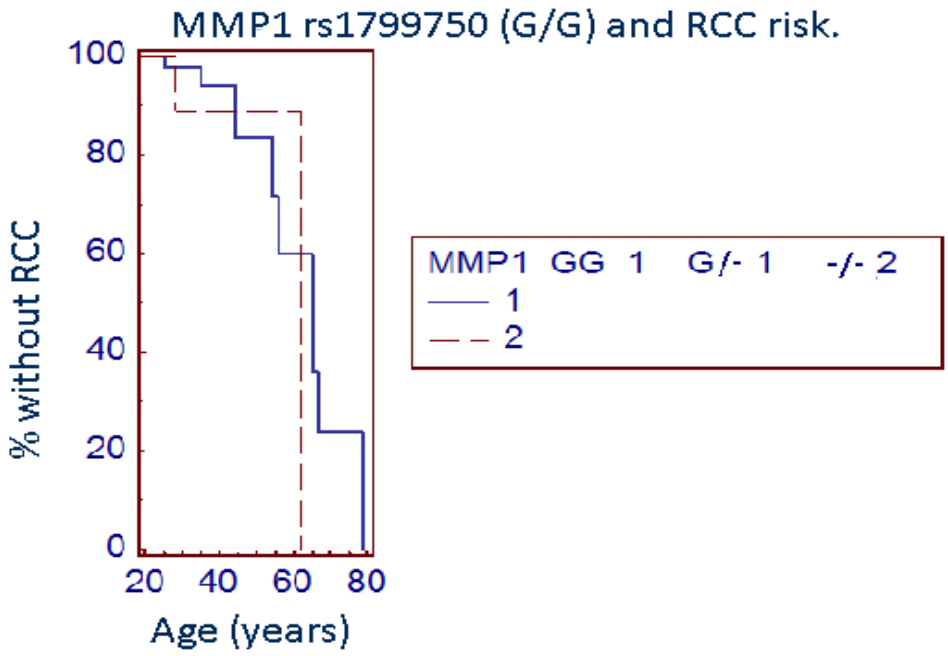
	Factors			
	1		2	
Sample size	58		6	
Median survival	62		67	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
25	0.981	0.0190	-	-
26	-	-	-	-
28	0.960	0.0275	-	-
30	-	-	-	-
32	-	-	-	-
34	-	-	-	-
35	0.928	0.0412	-	-
38	-	-	-	-
39	-	-	-	-
40	-	-	-	-
44	0.857	0.0784	-	-
45	-	-	-	-
47	-	-	-	-
54	0.734	0.132	-	-
56	0.612	0.157	-	-
62	0.490	0.166	-	-
65	0.245	0.148	-	-
67	-	-	0.000	0.000
73	-	-	-	-
79	0.000	0.000	-	-
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	10.0		1.0	
Expected n	9.5		1.5	
Chi-square	0.2146			
DF	1			
Significance	P = 0.6432			

Table 17 shows a P-value of 0.6432 - these results are not considered statistically significant.

MMP1 rs1799750 (G/G)

The polymorphic variation at *MMP1* is either a G allele or a deletion.

Figure 12 Survival Curve for *MMP1 rs1799750* and age of onset for RCC.



The survival curve in Figure 12 shows the G/G and G/- genotypes to be the least deleterious with the latest age of onset, the homozygous deletion (2) results in an age of onset for RCC at age 60.

Table 18 Statistical analysis of Figure 12

	Factors			
	1		2	
Sample size	50		13	
Median survival	65		62	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	-	-	-	-
22	-	-	-	-
24	-	-	-	-

	Factors			
	1		2	
Sample size	50		13	
Median survival	65		62	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
25	0.979	0.0210	-	-
26	-	-	-	-
28	-	-	0.889	0.105
30	-	-	-	-
32	-	-	-	-
34	-	-	-	-
35	0.942	0.0409	-	-
38	-	-	-	-
39	-	-	-	-
40	-	-	-	-
44	0.838	0.105	-	-
45	-	-	-	-
47	-	-	-	-
54	0.718	0.143	-	-
56	0.598	0.162	-	-
62	-	-	0.000	0.000
65	0.359	0.163	-	-
67	0.239	0.146	-	-
73	-	-	-	-
79	0.000	0.000	-	-
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	9.0		2.0	
Expected n	9.7		1.3	
Chi-square	0.5523			
DF	1			
Significance	P = 0.4574			

These results do not appear to have statistical significance based on the P-value in Table 18 which is quoted at 0.4574 above the P = >0.05 threshold.

Figure 13 Survival Curve for *MMP1 rs1799750* and age of onset for RCC.

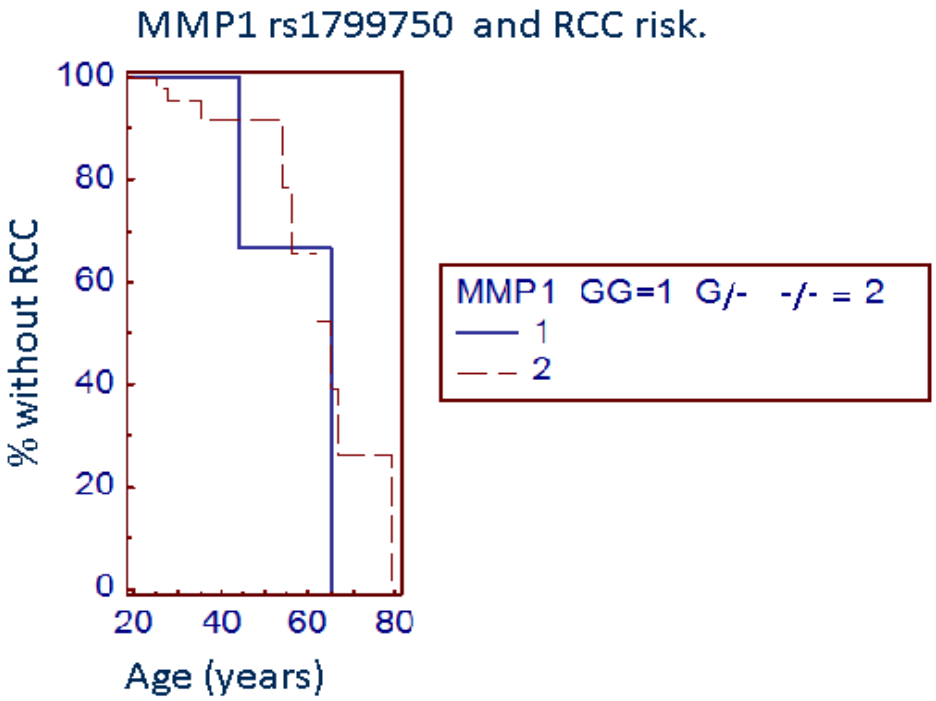


Figure 13:These results appear to be statistically insignificant given the P-value quoted in Table 19.

Table 19 Statistical analysis of Figure 13

	Factors			
	1		2	
Sample size	12		51	
Median survival	65		65	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	-	-	-	-
22	-	-	-	-
24	-	-	-	-
25	-	-	0.979	0.0206
26	-	-	-	-
28	-	-	0.956	0.0306

	Factors			
	1		2	
Sample size	12		51	
Median survival	65		65	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
30	-	-	-	-
32	-	-	-	-
34	-	-	-	-
35	-	-	0.919	0.0465
38	-	-	-	-
39	-	-	-	-
40	-	-	-	-
44	0.667	0.272	-	-
45	-	-	-	-
47	-	-	-	-
54	-	-	0.788	0.128
56	-	-	0.656	0.160
62	-	-	0.525	0.174
65	0.000	0.000	0.394	0.173
67	-	-	0.263	0.157
73	-	-	-	-
79	-	-	0.000	0.000
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	2.0		9.0	
Expected n	1.6		9.4	
Chi-square	0.1526			
DF	1			
Significance	P = 0.6961			

Hazard ratio	0.7496
95% CI	0.1384 to 4.0589

The P-value seen in Table 19 (0.6961) is above the threshold for statistical significance therefore these results are unlikely to be statistically significant.

Figure 14 Survival Curve for *MMP1 rs1799750* and age of onset for FF.

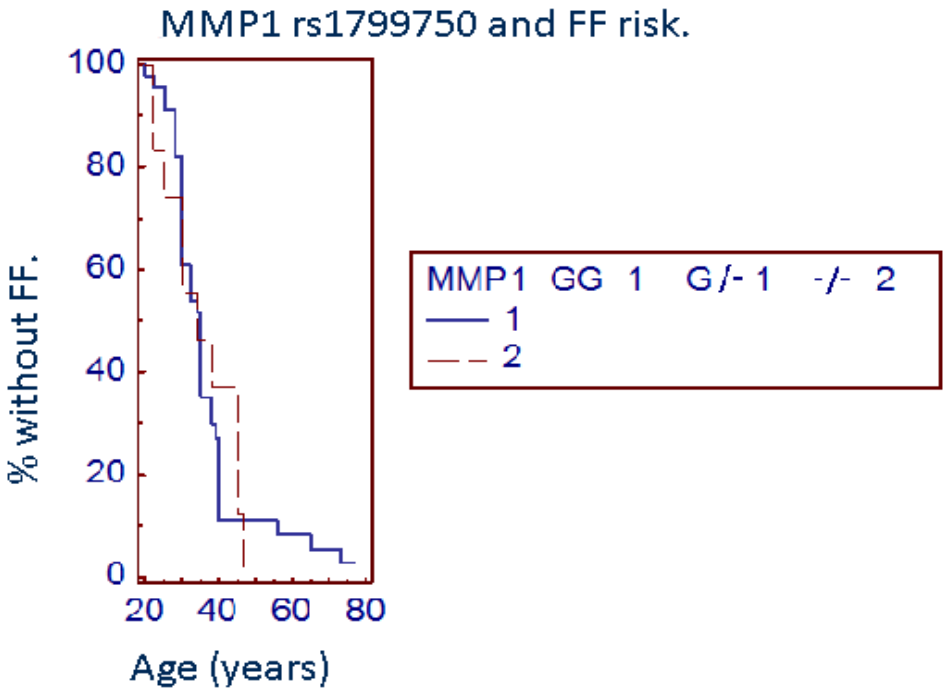


Table 20 Statistical analysis of Figure 13

	Factors			
	1		2	
Sample size	48		12	
Median survival	35		34	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	0.979	0.0206	-	-
22	0.958	0.0288	0.833	0.108
24	-	-	-	-
25	0.915	0.0408	0.741	0.129

	Factors			
	1		2	
Sample size	48		12	
Median survival	35		34	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error
26	-	-	-	-
28	0.821	0.0576	-	-
30	0.610	0.0742	0.556	0.149
32	0.539	0.0760	-	-
34	0.516	0.0762	0.463	0.150
35	0.352	0.0730	-	-
38	0.298	0.0711	0.370	0.146
39	0.271	0.0696	-	-
40	0.108	0.0503	-	-
45	-	-	0.123	0.112
47	-	-	0.000	0.000
56	0.0812	0.0444	-	-
65	0.0541	0.0370	-	-
73	0.0271	0.0266	-	-
77	-	-	-	-
	Comparison of survival curves (Logrank test)			
Endpoint: Observed n	40.0		10.0	
Expected n	39.8		10.2	
Chi-square	0.005666			
DF	1			
Significance	P = 0.9400			

These findings appear to be statistically insignificant given the P-value (0.94) quoted in Table 20.

Figure 15 Survival Curve for *MMP1 rs1799750* and age of onset for FF.

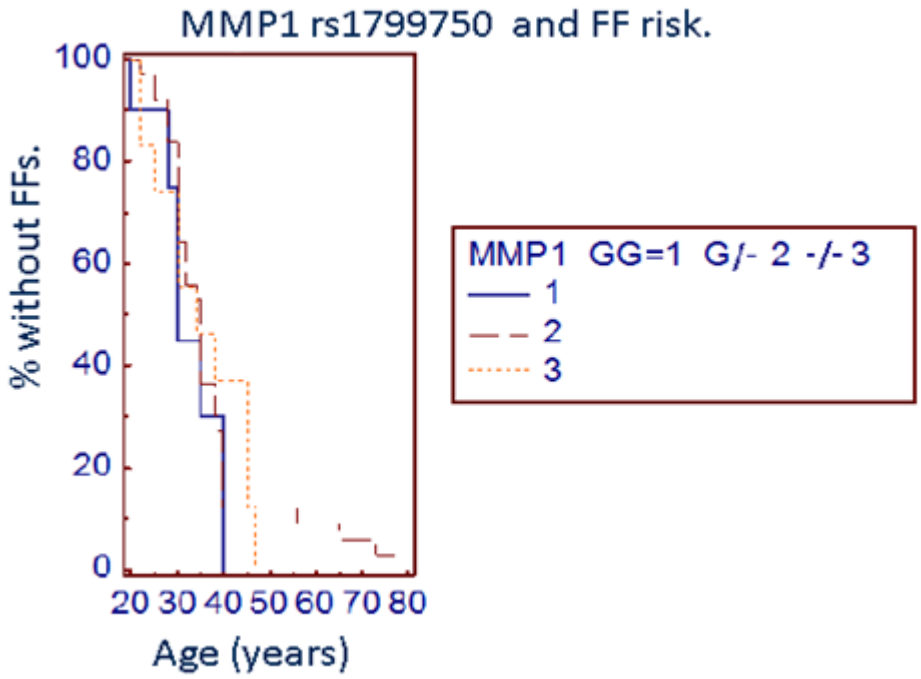


Table 21 Statistical analysis of Figure 14

	Factors					
	1		2		3	
Sample size	10		38		12	
Median survival	30		35		34	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error	Survival Proportion	Standard Error
20	0.900	0.0949	-	-	-	-
22	-	-	0.974	0.0260	0.833	0.108
24	-	-	-	-	-	-
25	-	-	0.921	0.0437	0.741	0.129
26	-	-	-	-	-	-
28	0.750	0.158	0.837	0.0609	-	-
30	0.450	0.190	0.642	0.0797	0.556	0.149
32	-	-	0.558	0.0827	-	-
34	-	-	0.530	0.0831	0.463	0.150

	Factors					
	1		2		3	
Sample size	10		38		12	
Median survival	30		35		34	
Survival time	Survival Proportion	Standard Error	Survival Proportion	Standard Error	Survival Proportion	Standard Error
35	0.300	0.176	0.363	0.0802	-	-
38	-	-	0.302	0.0774	0.370	0.146
39	-	-	0.272	0.0753	-	-
40	0.000	0.000	0.121	0.0562	-	-
45	-	-	-	-	0.123	0.112
47	-	-	-	-	0.000	0.000
56	-	-	0.0907	0.0496	-	-
65	-	-	0.0605	0.0413	-	-
73	-	-	0.0302	0.0297	-	-
77	-	-	-	-	-	-
	Comparison of survival curves (Logrank test)					
Endpoint: Observed n	6.0		34.0		10.0	
Expected n	4.7		35.1		10.2	
Chi-square	0.3724					
DF	2					
Significance	P = 0.8301					

Given the P-value seen in Table 21 it appears these results do not represent a statistically significant finding.

MC1R rs1805007 [T] and *ASIP rs1015362*[G]

Whilst these variants were screened within the patient samples, it was later revealed (due to inconsistencies in information on the target sequences) that the primers designed bound to a different region. Therefore these results do not represent the intended polymorphisms and are not included for analysis or discussion.

Analysis of Exome Sequencing in a renal tumour from a BHD patient

Blood and renal tumour DNA from a BHD patient with a history of two gastrointestinal tract tumours was analysed using exome sequencing. Although additional clinical data was not available, the tumour exome results were analysed to identify candidate loss of function mutations in candidate human cancer genes. Potentially interesting mutations were selected and prioritised based upon mutation type and relevance to BHD and renal cancer. The criteria used from high to low were:

Somatic Nonsense mutation, Somatic Frameshift mutation, Splice site mutation predicted to affect splicing, missense substitution predicted to be pathogenic and linked to cancer, somatic substitution predicted to be pathogenic but not linked to cancer and finally benign substitution predicted to be non-pathogenic and not linked to cancer. Using these criteria the mutations most likely to be deleterious and associated with a significant increase to cancer vulnerability were selected.

After this these mutations and genes were run through COSMIC software alongside manual selection to determine their relevance to cancer, in particular relevance to kidney cancers such as VHL or *FLCN*. I then undertook mutation analysis of the selected variations in sporadic RCC to help determine if the gene of interest was likely to be a frequent target of somatic mutation in renal tumourigenesis. Ultimately this approach might identify genes implicated in BHD and sporadic renal tumourigenesis, and germline variants that affected the function of such genes could then be investigated as candidate modifier alleles for RCC risk in BHD patients.

FLCN:

Within the exome data, the first patient presented a c.G907A:p.G303R non-synonymous mutation (NM_144606) in *FLCN* on exon 8. This germline mutation did not appear on the *FLCN* mutation database and appeared to be novel (Leiden University Medical Center, 2009).

The second tumour that underwent exome analysis came from a Japanese patient who presented with a homozygous/hemizygous splicing variation on exon 10 - c.1062+6C>T on *FLCN* and two separate intronic changes. This splice-site mutation has previously been categorised as a SNP (rs8065832) in a study attempting to find a link between Folliculin mutations and severe COPD (Chronic obstructive pulmonary disease however, no significant association was found between them (Choo, et al., 2008).

TP53:

This gene is the most commonly mutated tumour suppressor gene in human cancer. Analysis of the tumour exome data revealed an exon 4 mutation c.C215G:pP72R. This mutation has previously been reported to be pathogenic. Variants at this location including c.G215C:pR72P have a demonstrated association with survival rates in advanced ovarian cancer syndromes causing functional changes (Wang, et al., 2004) and differences in apoptotic potential (Dumont, et al., 2003). Though *TP53* mutations are reported in kidney cancer, inspection of the exome sequencing results for blood DNA revealed that the *TP53* mutation was also present in blood and was not a somatic mutation.

The British patient presented with multiple mutations within *TP53* as described below:

Table 22 *TP53* variants detected

Exon	**Type**	Status
Exon 5	c.T526C:p.C176R	Novel
Exon 1	c.T130C:p.C44R	Novel
Exon 4	c.C215G:pP72R	Confirmed Pathogenic

The tumour from the Japanese patient also presented with the c.C215G:p72R variant on *TP53.* This change is well documented and has been associated with multiple cancer syndromes and is categorised as the Rs1042522 variant (Johnson, et al., 2007).

ARHGEF5

The first potential modifier gene selected was *ARHGEF5* located on 7q35 also known as Rho guanine nucleotide exchange factor which produces a protein of the same name. This oncogene is responsible for producing proteins stimulated by Rho-factors which have critical roles in extracellular communication and signalling and are involved in organisation of the cytoskeleton; both of these attributes are found dysregulated within cancerous growths and contribute to progression and metastasis (Debily, et al., 2004). Evidence suggests *ARHGEF5* uses RhoA factors to regulate dendritic cell migration, and has been shown to significantly alter expression of RhoA and RhoB in kidney cells; RhoC and G are less affected by this (Wang, et al., 2009). *ARHGEF5* has been implicated in breast cancer, and it is currently known that Folliculin interacts and regulates with Rhoa-A and rho-guanine factors leading to altered RhoA signalling and cell adhesion (Nahorski, et al., 2012) (Medvetz, et al., 2012). The patient in our exome study had a heterozygous "stopgain" or nonsense mutation on exon 2

with no other novel changes. The Japanese patient's tumour presented with three intronic variations on *ARHGEF5* but no exonic variations.

Znf180

ZNF180 encodes Zinc Finger Protein 180 and is 5 exons long, Zinc fingers are involved in up/down regulation and gene expression, abnormalities typically result in a deficiency in DNA repair mechanisms which can naturally lead towards malignancies (Wang, et al., 1996). The patient presented with a novel heterozygous frameshift mutation on exon 5 (c.471delT:p.D157fs).

IGFBP2

Insulin-Like Growth Factor Binding protein 2 (*IGFBP2*) is produced by the gene of the same name located on chromosome 2q35; it contains 4 exons and 3 introns. These growth factors modulate growth and development through inhibition or stimulation using IGF receptors, and have been linked to breast cancers through endothelial recruitment and signalling (Png, et al., 2011). There is already a precedent for *IGFBP2* over-expression to result in increased resistance to chemotherapy and its role as a survival factor in breast cancer cell proliferation (Foulstone, et al., 2013). *IGFBP2* factors are implicated in renal diseases through over-expression of these growth factors (Wolf, et al., 2000). Levels of *IGFBP2* expression are found to be increased in patients with active prostate carcinoma (Jean Ho & Baxter, 1997). A recent study determined the *IGFBP2* to be a predictor of renal function deterioration in type II diabetes where an increase in *IGFBP2* expression resulted in a greater mortality and faster decline in kidney function (Narayanan, et al., 2012). The patient presented with a novel heterozygous non-frameshift insertion of "CGC" at position 64-65.

FOXC1

The *FOXC1* or "Forkhead Box C1" gene is located a chromosome 6p25; this gene family is named after the "fork-headed" DNA-binding domain and their role as transcription factors. As a result *FOX* regulates gene expression and is associated with developmental roles during embryogenesis; their regulation affects proliferation and differentiation of cells and cell growth making them potential targets for mutation during carcinogenesis. Over-expression of the *FOXC1* proteins result in reduced patient survival in lung cancer (Wei, et al., 2013) and pancreatic ductal adenocarcinoma (Wang, et al., 2013), hepatocellular carcinoma (HCC) through tumour metastasis (Xia, et al., 2013) and microvascular invasion in HCC (Xu, et al., 2012). *FOXC1* is necessary for the formation of the kidney and urinary tract (Kume, et al., 2000) with mutations in *FOXC1* resulting in congenital abnormalities in human kidneys (Nakano, et al., 2003). The severe patient presented with a novel heterozygous non-frameshift insertion of GCG at position 1141-1142. The Japanese tumour sample presented with a nonframeshift insertion of three bases (c.1361_1362insCGG:p.G454delinsGG).

Of the four genes selected as potential genes mutated in sporadic RCC, *ZNF180, IGFBP2* and *FOXC1* were unable to be fully screened due to time constraints. The *ARHGEF5* gene was fully completed and showed multiple novel variations listed below:

Analysis of candidate genes in sporadic renal tumours:

Sequencing analysis of *ARHGEF5* in tumour DNA presented several novel sequence variations not found within current literature. These results were obtained from a sample size of 37 patients with sporadic RCC.

Table 23 *ARHGEF5* novel variations.

Variant	Location	Exon	Polyphen/Sift
Stopgain	C.29insGCTGA p.Ala29GlySer30*	2	Truncating
Synonymous	c.1283C>T (p(=))	2	Benign
Synonymous	c.1368C>A (P(=))	2	Benign
Non-Synonymous	c.1845G>C p.Gln615Hist	2	Damaging
Stop Codon	c.2870C>G p.S957*	2	Truncating
Non-synonymous	c.2933G>T p.Ser978Ile	3	Possibly damaging
Synonymous	c.3136T>C (p(=))	3	Benign
Non-synonymous	c.3409T>C p.Ser1137Pro	6	Benign

Any novel variations detected were analysed using Polyphen and Sift software with the aim to predict any functional effects these variations might cause. The majority of these variants appeared to be benign with a small number of them predicted to be damaging.

Two clear truncating mutations were detected in tumour DNA. A nonsense mutation at codon 29 in exon 2 was detected in the exome data from the tumour DNA of the BHD patient and confirmed by Sanger sequencing. In addition a second c.2870C>G p.S957* mutation at codon 957 was detected in one of the 37 sporadic RCC tested. Neither of these changes were detected in the matched "normal" (non-tumour) DNA samples for both these patients and so this was consistent with a somatic mutation.

The non-synonymous substitution of glutamine (GLN) to glutamic acid (GLU) was detected in another patient but was not found within their normal DNA sample suggesting a somatic change. Polyphen/SIFT predict this variation to be benign.

The predicted damaging mutations were also confirmed to be somatic changes in the patients that presented with them; none of the damaging mutations were shared between patients. With the exception of the truncating variant, which we originally detected in the tumour from the British patient with BHD syndrome. Out of the 37 patients screened only 3 novel damaging variants were detected in three separate cases.

Discussion:

With the aim of identifying possible alleles that might modify or predispose a BHD patient with a Folliculin mutation towards a specific phenotype such as melanoma or renal cell carcinomas; sequencing and statistical analysis was performed on approximately 70 BHD patients.

Of the polymorphisms analysed; only *MITF* showed a lack of any variation within the 72 samples with every patient displaying a G/G genotype. This was not unexpected given the noted low frequency of the deleterious A allele within a population (Yokoyama, et al., 2011) and the relatively small sample size used in this study. Yokoyama reported the *MITF* polymorphism with a frequency of 0.0085 in controls and 0.0176 within patients with a melanoma within a UK population with only 52 patients with the variant and 3992 without. The Australian sample group showed a similar ratio but with a lower frequency of the allele than the UK. With the Australian and UK frequencies combined the overall frequency for the allele was 0.0171 in affected cases and 0.0079 in control groups. Given these circumstances the lack of any variation is less surprising; it may still be a potential modifier for melanoma within BHD but this study simply lacked enough samples to for such a low frequency variant to be present within the patient samples. A more in-depth study with a wider sample size may reveal a possible predisposition to the associated skin phenotypes to BHD. Had there been variation within this polymorphism especially with such a small sample size it might suggest a strong link between a deleterious *MITF* variant and fibrofolliculoma risk.

Within the literature the T allele of *ASIP* rs4911414 has been shown to associate with a higher risk of melanoma (Kanetsky, et al., 2002). Analysis of the *ASIP* rs4911414 variant shows statistically insignificant results in association with fibrofolliculoma risks with P-values

of greater than 0.05 (P =0.9916 and P =0.9788) these figures are far outside a significance threshold and cannot be relied upon.

The 11q13.3 locus variant rs7105934 only presented with the G/G and G/A haplotypes within this study, there were no homozygote A/As. The A allele has been associated in previous studies with a protective effect against renal cell carcinoma with an OR of 0.69, however evidence suggests environmental and lifestyle factors can result in the loss of this protective effect. Factors such as smoking, obesity and heavy drinking have been implicated (Purdue, et al., 2011). Given the statistical insignificance of the findings (P = 0.4247 and 0.3480) presented it is less likely that this polymorphism modifies RCC risk.

When compared with age of onset for fibrofolliculomas, *TYR rs1126809 (R402Q)* the G/G haplotypes in both Figure 7 and 8 are associated with the earliest age of onset. In figure 8 the A/A homozygote results in a later onset than G/G but earlier than G/A. In this study only three patients presented with an A/A homozygote out of 67 samples with G/G having 28 patients and G/A 30. This might partially explain the abnormal curve for the A/A in comparison with the other two haplotypes, though it does not account for why the G/G haplotype has the earliest age of onset. In the genome-wide association study of approximately 4000 patients diagnosed with basal cell carcinomas or cutaneous melanomas by (Gudbjartsson, et al., 2008) the R402Q variant was shown to increase the risk of cutaneous melanoma with an odds ratio of 1.21 alongside basal cell carcinomas with an odds ratio of 1.14. These findings suggest both A/A and G/A genotypes confer increased risk and contrast with the findings of the prior studies. Nevertheless the findings in my patient group were statistically significant. There are several explanations for this discrepancy. Firstly the mechanisms of tumourigenesis in fibrofolliculomas in BHD patients may differ

from those in sporadic melanoma and an allele protective for melanoma might predispose to fibrofolliculomas. The tumourigenesis processes in fibrofolliculomas may differ from those in melanoma and basal cell carcinoma such that different modifier genes are implicated (though ultra violet radiation exposure is a risk factor for both melanoma and basal cell carcinoma it has not been implicated in fibrofolliculomas). Little is known about tumourigenic processes in fibrofolliculomas in BHD syndrome and how these compare to melanoma and so it is not possible to refute or confirm this hypothesis. A second possibility is the small sample size (<70) in my study resulted in a false positive finding. Although the effect of the *TYR* SNP appeared statistically significant, this was without correcting for multiple testing and application of Bonferroni correction which might have revealed this result to be a false positive. Nevertheless this finding does suggest that the SNP should be analysed in a larger group of patients.

EPAS rs11894252 was tested against age of onset for renal cell carcinoma. Findings show statistically insignificant P-values of 0.8922 and 0.6432 suggesting these findings are not associated with RCC risk.

For the *MMP1 rs1799750* polymorphism G/G variant in the literature is associated with an increased risk of several cancers (Ricketts, et al., 2009). However, this study lacks the statistical power to support these findings completely given the P values in Table 17 and 18 (P = 0.6961 and 0.94 respectively). Within the patient samples, the majority of the patients developed fibrofolliculomas with very few developing RCC it could be argued this polymorphism may also increases the likelihood of developing FFs however statistical evidence does not support this hypothesis given the insignificant P-value of 0.8301 seen in Table 19 suggesting there is no correlation.

ARHGEF5:

In terms of the exome project, the stop-codon insertion at the start of exon 2 on *ARHGEF5* showed promise, appearing both in the BHD patient with the severe phenotype and a somatic *p53* mutation and one of our sporadic RCC patients. This variation appeared in neither patient's normal DNA samples and only appeared within tumour derived DNA samples. The addition of a stop-codon results in truncation of the transcript for exon 2 and results in a non-functional protein which would most likely have an effect on the phenotype. Given the role of *ARHGEF5* in cell structure organisation and extracellular communication/RhoA signalling in Kidney cells, this gene represents a likely target for a modifier of renal carcinoma risk within a BHD patient. However, out of the 37 patients only one other patient presented with this somatic change. It is difficult to determine whether this variation would be a driving factor within renal cancer or whether it might represent a "passenger mutation" (Muller, et al., 2012). The gene is associated with aggressive tumourigenesis and metastasis within breast cancers (Debily, et al., 2004) and may also provide a similar role in renal cell carcinomas. This particular variation was currently unreported in literature or websites documenting variation so this change is reasoned to be a novel variant. The two other variations predicted to be damaging were present individually in two separate patients' tumour DNA samples both variants are non-synonymous changes resulting in a different amino acid being produced by the transcript. These changes were not found in any tumours or current literature, Polyphen/Sift software predicts them as damaging/possibly damaging. The c.2870C>G p.S957* stopgain mutation has been documented on COSMIC (COSM1488330) and has previously been found within a breast cancer sample (Wellcome Sanger Trust Institute, 2013). Further investigation would

be required to determine the effects and implications of these possible mutations in tumourigenesis including a larger sample size of patients.

Finally several synonymous or "silent" novel variants within the *ARHGEF5* gene were found within patient tumour DNA. These variants are a change in the sequence but given the redundancy of the base combinations used in producing specific amino acids these variants do not change the amino acid or the final protein product. Polyphen/SIFT software describes them as benign and unlikely to have any effect on phenotype as they do not impair or modify protein production. However it is worth noting that variants which are classified as silent may affect splicing sites for other proteins produced by the same transcript. Ultimately though synonymous changes are a low priority and any variations or mutations that might modify the risk or progression of tumourigenesis are likely to be non-synonymous or stopgain mutations.

Study Limitations:

Limitations of analysing SNPs as modifiers of BHD phenotype risk:

Given the nature of a rare disease such as BHD, this particular study suffers from a small sample size of patients making it difficult to draw any decisive conclusions towards modifying risk of a specific phenotype. Without germline DNA from very large numbers of BHD patients and with incomplete clinical data it would be difficult to detect subtle modifying effects that a common polymorphic variant might exert. One of the main arguments behind SNP polymorphisms driving a specific phenotype is that they might modify a particular outcome by a very small percentage; often they work in tandem with other SNPs and modify risk more than they would individually. Naturally this makes detecting the effect of a SNP alone very difficult without significant sample sizes such as a genome-wide association study (GWAS). GWAS often have sample sizes within the thousands and often across different populations should there be a noticeable difference in allele frequency within a given population. Given our small sample size the lack of a positive finding does not necessarily indicate that a SNP does *not* modify a particular outcome in BHD merely that we lack the numbers and statistical power to prove that it does.

The strength of a Kaplan-Meier analysis is its effectiveness in processing censored patient data and allows for the inclusion of incomplete data for example if a patient has withdrawn from a study or their symptoms are no longer reported. The logrank test was an appropriate nonparametric (lack of parameters or structure assumptions within the data) statistical study to analyse data generated by a Kaplain-Meier curve given its tolerance for right-censored data. These tests allowed for a greater statistical significance to be gleaned from the data, however with one exception (*TYR*) these tests were not enough. The majority of

results were statistically non-significant and we must accept the null hypothesis. It's important to avoid over-interpreting a single statistically significant result within a series of statistically insignificant results; it is possible a result like this might occur through random chance alone (Bland & Altman, 1995). Bonferroni correction is a conservative method which aims to prevent occurrences of this arising from multiple tests within the same subset of data. This method can be used to avoid Type I statistical errors (rejecting a null hypothesis based on a false positive result) it is less effective and may increase the risk of a Type II error (when an incorrect null hypothesis is not rejected) (Genomics and Bioinformatics Group, 2013).

Another potential flaw within the study especially given the recurring problems of a limited sample size is the presence of family members or large families within the data. Related patients might skew data towards a certain outcome, as they would likely share similar polymorphisms and inherited genotypes causing a greater representation of a particular genotype than you might expect from a random population. Whilst the data included a family of five people, since censored data does not take this factor into account it might exacerbate issues with false positives. The SNP in *TYR* was statistically significant (without correction for multiple testing), however this study found opposite and contradictory findings to previous literature so this result may be a false finding.

Ultimately a larger study would be required to detect subtle effects that may be caused by the suggested SNPs; it may also be informative to use multiple different populations within the study. Several SNPS are noted to vary in frequency such as the *MC1R* polymorphisms and any effects they might have would be lost in a study using an inappropriate population set. *EPAS1 rs11894252* is an excellent example of this variation as the minor allele within a

global population is reversed in a European population becoming the most common variant as seen in Table 7. In summary, the work reported here can provide the basis for a larger follow up study. Though I have been cautious interpreting the importance and validity of my positive finding it would be important to follow up the finding and confirm or refute its significance.

Limitations of analysing candidate genes in sporadic renal tumours:

A major limitation of this study was the available time, as a result several promising candidate genes were selected with primers designed and tested but given time constraints were unable to be sequenced in all of the tumours. The short amount of time available was exacerbated by difficulty producing reliable primers for each region as they had to be designed manually as well as availability of PCR machines and the sequencer. These problems were compounded by mechanical failures resulting in the need to replace or repair these machines in order to progress and generate additional data. Whilst these problems are out of the scope of the study's design they were a contributing factor towards the amount of data produced. Given the size of *ARHGEF5* testing all 15 exons within 37 patients took a fair amount of time to accomplish.

Whilst the limitations of the use of exome sequencing described previously have been raised it is still important to confirm information used from it to avoid false positive results. This study tested the severe patient's normal and tumour DNAs for mutations in the candidate genes that were selected to confirm their presence before analysing other patient DNA samples. This was an effort to avoid searching for variations that weren't there based off misinformation generated by large-scale testing.

It was interesting that a *TP53* mutation detected by exome sequencing in the tumour DNA from a BHD patient was present in the germline and was not a somatic mutation. This patient had an unusual phenotype for a BHD patient (colorectal, oesophageal and renal cancers) and it can be suggested that this severe phenotype might have resulted from the combination of germline *FLCN* and *TP53* mutations. It also raises the possibility that functional genetic variants in *TP53* might be tested as candidate modifiers in BHD patients.

Though *ARHGEF5* was not commonly mutated in sporadic RCC it is interesting to note that at least one somatic truncating mutation was detected in sporadic RCC and it would be interesting to undertake a larger study comprising more BHD and sporadic tumours for comparison.

List of References:

Works Cited

Allwood, J. S. & Harbinson, S., 2013. SNP model development for the prediction of eye colour in New Zealand. *Forensic Science International: Genetics,* pp. 1443-1452.

Applied Biosystems, 2010. *BigDye® Terminator v3.1 Cycle Sequencing Kit.* [Online] Available at: http://tools.invitrogen.com/content/sfs/manuals/cms_041329.pdf [Accessed 1 July 2013].

Artandi, S. E. & DePinho, R. A., 2010. Telomeres and telomerase in cancer. *Carcinogenesis,* Volume 31, pp. 9-18.

Baker, D. J. et al., 2011. Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. *Nature,* Volume 479, pp. 232-236.

Banno, K. et al., 2012. Epimutation and cancer: a new carcinogenic mechanism of Lynch syndrome (Review). *International Journal of Oncology,* Volume 41, pp. 793-797.

Bardessy, N., Beckwith, J. B. & Pelletier, J., 1995. Clonal Expansion and Attenuated Apoptosis in Wilm's Tumours Are Associated with p53 Gene Mutations. *Cancer Research,* Volume 55, pp. 215-219.

Bastola, P. et al., 2013. Folluclin contributes to VHL tumour suppressing activity in renal cancer through regulation of autophagy.. *PLoS One,* 8(7).

Bastola, P. et al., 2013. Folliculin Contributes to VHL Tumour Suppressing Activity in Renal Cancer through Regulation of Autophagy. *PLOS One,* 8(7).

Baylin, S. B. & Ohm, J. E., 2006. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction?. *Nature Reviews Cancer,* Volume 6, pp. 107-116.

Bielinksa, B. et al., 2013. Association of the BRCA1 promoter polymorphism rs11655505 with the risk of familial breast and/or ovarian cancer.. *Familial Cancer.*

Bishop, D. T. et al., 2009. Genome-wide association study identifies three loci associated with melanoma risk.. *Nature Genetics,* 41(8), pp. 920-925.

Bissig, H. et al., 1999. Evaluation of the Clonal Relationship between Primary and Metastatic Renal Cell Carcinoma by Comparative Genomic Hybridization.. *American Journal of Pathology,* 155(1), pp. 267-274.

Bland, J. M. & Altman, D. G., 1995. Multiple significance tests: the Bonferroni method. *British Medical Journal,* Volume 310, p. 170.

Bliss, J. M. et al., 1995. Risk of cutaneous melanoma associated with pigmentation characteristics and freckling: systematic overview of 10 case-control studies. The international melanoma Analysis Group (IMAGE).. *International Journal of Cancer,* 62(4), pp. 367-376.

Bourdon, J. C., 2007. p53 and its isoforms in cancer. *British Journal of Cancer,* 97(3), pp. 277-282.

Box, N. F. et al., 2001. MC1R genotypes modifies risk of melanoma in families segregating CDKN2A mutations.. *American Journal of Human Genetics,* 69(4), pp. 765-773.

Brogna, S. & Wen, J., 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology,* Volume 16, pp. 107-113.

Bunz, F. et al., 1998. Requirement for p53 and p21 to sustain G2 arrest after DNA damage.. *Science,* 282(5393), pp. 1497-1501.

Byrum, S. D. et al., 2013. Quantitative Proteomics Identifies Activation of Hallmark Pathways of Cancer in Patient Melanoma. *Journal of Proteomics & Bioinformatics,* 6(3), pp. 43-50.

Cao, Q. et al., 2012. Chromosome 11q13.3 variant modifies renal cell cancer risk in a Chinese population. *Mutagenesis,* 27(3), pp. 345-350.

Cardeñosa, E. E. et al., 2012. Low penetrance alleles as risk modifiers in familial and sporadic breast cancer.. *Familial Cancer,* 11(4), pp. 629-636.

Chatzinasiou, F. et al., 2011. Comprehensive Field Synopsis and Systematic Meta-Analysis of Genetic Association Studies in Cutaneous Melanoma. *Journal of the National Cancer Institute,* Volume 103, pp. 1-9.

Cheng, N. et al., 2008. TGF-β signaling deficient fibroblasts enhance Hepatocyte Growth Factor signaling in mammary carcinoma cells to promote scattering and invasion. *Molecular Cancer Research,* 6(10), pp. 1521-1533.

Chen, Y. & Mattey, D., 2012. Age at onset of rheumatoid arthritis: association with polymorphisms in the vascular endothelial growth factor A (VEGFA) gene and an intergenic locus between matrix metalloproteinase (MMP) 1 and 3 genes. *Clinical and Experimental Rheumatology,* 30(6), pp. 894-898.

Choi, M. et al., 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Science USA,* 106(45), pp. 19096-19101.

Choo, M. H. et al., 2008. Folliculin mutations are not associated with severe COPD. *BMC Medical Genetics,* 9(120).

Cocciolone, R. A. et al., 2010. Multiple Desmoplastic Melanomas in Birt-Hogg-Dubé Syndrome and a Proposed Signaling Link Between Folliculin, the mTOR Pathway, and Melanoma Susceptibility. *Archives of Dermatology,* 146(11), pp. 1316-1318.

Cronin, J. C. et al., 2009. Frequent mutations in the MITF pathway in melanoma. *Pigment Cell Melanoma Research,* 22(4), pp. 435-44.

Dameron, K. M., Volpert, O. V., Tainsky, M. A. & Bouck, N., 1994. Control of angiogenesis in fibroblasts by p53 regulation of thrombospondin-1. *Science,* 265(5178), pp. 1582-1584.

Daniali, L. et al., 2013. Telemeres shorten at equivalent rates in somatic tissues of adults.. *Nature Communications,* Volume 4.

Debily, M.-A.et al., 2004. Expression and molecular characterization of alternative transcripts of the ARHGEF5/TIM oncogene specific for human breast cancer.. *Human Molecular Genetics,* 13(3), pp. 323-334.

Dumont, P. et al., 2003. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. *Nature Genetics,* Volume 33, pp. 357-365.

Fatemi, M. J. et al., 2005. Footprinting of mammalian promoters: use of CpG DNA methyltransferase revealing nucleosome posiiton at a single molecule level. *Nucleic Acids Research,* 33(20), p. 176.

Fernandes da Silva, N. et al., 2003. Analysis of the Birt–Hogg–Dubé (BHD) tumour suppressor gene in sporadic renal cell carcinoma and colorectal cancer. *Journal of Medical Genetics,* 40(11), pp. 820-824.

Foley, C. J. et al., 2013. Matrix metalloprotease 1a deficiency suppresses tumor growth and angiogenesis. *Oncogene.*

Foster, K. et al., 1994. Somatic mutations of the von Hippel-Lindaudisease tumour supressor gene in non-familial clear cell renal carcnoma. *Human Molecular Genetics,* 3(12), pp. 2169-2173.

Foulstone, E. J., Zeng, L., Perks, C. M. & Holly, J. M., 2013. Insulin-like growth factor binding protein 2 (IGFBP-2) promotes growth and survival of breast epithelial cells: novel regulation of the estrogen receptor.. *Endocrinology,* pp. 1780-1793.

Gad, S. et al., 2007. Mutations in BHD and TP53 genes, but not in HNF1beta gene, in a large series of sporadic chromophobe renal cell carcinoma. *British Journal of Cancer,* 96(2), pp. 336-340.

Gao, D. et al., 2008. Endothelial Progenitor Cells Control the Angiogenic Switch in Mouse Lung Metastasis. *Science,* Volume 319, pp. 195-198.

Garraway, L. A. et al., 2005. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature,* Volume 436, pp. 117-121.

Gaudet, M. M. et al., 2013. Identification of a BRCA2-specific Modifier Locus at 6p24 Related to Breast Cancer Risk. *PLOS Genetics,* 9(3).

Gaur, K. et al., 2013. The Birt-Hogg-Dube tumor suppressor Folliculin negatively regulates ribosomal RNA synthesis. *Human Molecular Genetics,* 22(2), pp. 284-299.

Genomics and Bioinformatics Group, 2013. *Microarray Data Analysis.* [Online] Available at: http://discover.nci.nih.gov/microarrayAnalysis/Statistical.Tests.jsp [Accessed 12 September 2013].

Ghilardi, G. et al., 2001. Matrix metalloproteinase-1 promoter polymorphism1G/2G is correlated with colorectal cancer invasiveness. *Clinical Cancer Research,* 7(8), pp. 2344-2346.

Gnarra, J. R. et al., 1994. Mutations of the VHL tumour-suppressor gene in Renal-Carcinoma. *Nature Genetics,* 7(1), pp. 85-90.

Gudbjartsson, D. F. et al., 2008. ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma.. *Nature Genetics,* 40(7), pp. 886-891.

Gudmundsson, J. et al., 2009. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nature Genetics,* 41(10), pp. 1122-1126.

Halaban, R. et al., 2000. Endoplasmic reticulum retention is a common defect associated with tyrosinase-negative albinism. *Proceedings of the National Academy of Sciences of the United States of America,* 97(11), pp. 5889-5894.

Hanahan, D. & Folkman, J., 1996. Patterns and Emerging Mechanisms of the Angiogenic Switch during Tumorigenesis. *Cell,* 86(3), pp. 353-364.

Hanahan, D. & Weinberg, R. A., 2000. The Hallmarks of Cancer. *Cell,* Volume 100, pp. 57-70.

Harding, R. M. et al., 2000. Evidence for Variable Selective pressures at MC1R. *The American Journal of Human Genetics,* 66(4), pp. 1351-1361.

Hayflick, L. & Moorhead, P., 1961. The serial cultiivation of human diploid cell strains.. *Experimental Cell Research,* pp. 585-621.

Herman, J. G. et al., 1994. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proceedings of the National Academy of Science,* Volume 91, pp. 9700-9704.

Hirata, H. et al., 2004. Association of a haplotype of matrix metalloproteinase (MMP1)-1 and MMP-3 polymorphisms with renal cell carcinoma. *Carcinogenesis,* 25(12), pp. 2379-2384.

Hughes, S. et al., 2007. Matrix Metalloproteinase Single-Nucleotide Polymorphisms and Haplotypes Predict Breast Cancer Progression. *Clinical Cancer Research,* 13(22), pp. 6674-6680.

Hutton, S. M. & Spritz, R. A., 2008. A Comprehensive Genetic Study of Autosomal Recessive Occular Albinism in Caucasian Patients. *Investigative Opthalmology & Visual Science,* 49(3), pp. 868-872.

Irizarry, R. A. et al., 2009. Genome-wide methylation analysis of human colon cancer reveals similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics,* 41(2), pp. 178-186.

Jean Ho, P. & Baxter, R. C., 1997. Insulin-like growth factor-binding protein-2 in patients with prostate carcinoma and benign prostatic hyperplasia. *Clinical Endocrinology,* 46(2), pp. 145-154.

Jerant, A. F., Johnson, J. T., Sheridan, C. D. & Caffrey, T. J., 2000. Early detection and treatment of skin cancer.. *American Family Physician,* 62(2), pp. 357-68.

Johnson, N. et al., 2007. Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility.. *Human Molecular Genetics,* 16(9), pp. 1051-7.

Kanamori, Y. et al., 1999. Correlation between Expression of the Matrix Metalloproteinase-1 Gene in Ovarian Cancers and an Insertion/Deletion Polymorphism in Its Promoter Region. *Cancer Research,* Volume 59, pp. 4225-4227.

Kanetsky, P. A. et al., 2002. A polymorphism in the agouti signaling protein gene is associatd with human pgimentation.. *American Journal of Human Genetics,* 70(3), pp. 770-775.

Khoo, S. K. et al., 2002. Clinical and genetic studies of Birt-Hogg-Dubé syndrome. *Journal of Medical Genetics,* Volume 39, pp. 906-912.

Khoo, S. K. et al., 2003. Inactivation of BHD in sporadic renal tumours. *Cancer Research,* 63(15), pp. 4583-4587.

Knudson, A. G., 1971. Mutations and Cancer: Statistical Study of Retinoblastoma. *Proceedings of the National Academy of Science of the United States of America,* 68(4), pp. 820-823.

Kozlowski, J. M., Hart, I. R., Fidler, I. J. & Hanna, N., 1984. A Human Melanoma Line Heterogeneous with Respect to Metastatic Capacity in Athymic Nude Mice. *Journal of the National Cancer Institute,* Volume 72, pp. 913-917.

Kume, T., Deng, K. & Hogan, B. L., 2000. Murine forkhead/winged helix genes Foxc1 (Mf1) and Foxc2 (mfh1) are required for the early organogenesis of the kidney and urinary tract.. *Development ,* 127(7), pp. 1387-95.

Kuukasjarvi, T. et al., 1997. Genetic Heterogeneity and Clonal Evolution Underlying Development of Asynchronous Metastasis in Human Breast Cancer. *Cancer Research,* Volume 57, pp. 1597-1604.

Lahmann, C., Bergemann, J., Harrison, G. & Young, A. R., 2001. Matrix metallioproteinase-1 and skin ageing in smokers. *The Lancet,* 357(9260), pp. 935-936.

Leiden University Medical Center, 2009. *Leiden Open Variation Database- The Folliculin Mutation Database.* [Online]
Available at: http://skingenedatabase.com/home.php?select_db=FLCN
[Accessed 12 September 2013].

Linehan, W. M., Walther, M. M. & Zbar, B., 2003. The genetic basis of cancer of the kidney. *Journal of Urology,* 170(6), pp. 2163-2172.

Lundgren, K. et al., 2008. Gene products of chromosome 11q and their association with CCND1 gene amplification and tamoxifen resistance in premenopausal breast cancer. *Breast Cancer Research,* p. 10.

Maina, E. N. et al., 2005. Identification of novel VHL target genes and relationship to hypoxia response pathways. *Oncogene,* Volume 24, pp. 4549-4558.

Makova, K. & Norton, H., 2005. Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans.. *Peptides,* 26(10), pp. 1901-1908.

Mazur, M. A. et al., 2013. Micropthalmia transcription factor (Mitf) regulates pancreatic β cell function. *Diabetes.*

McDougall, S. R., Anderson, A. R. & Chaplain, M. A., 2006. Mathematical modelling of dynamic adaptive tumour-induced angiogenesis: clinical implications and therapeutic targeting strategies.. *Journal of Theoretical Biology,* 241(3), pp. 564-89.

Medvetz, D. A. et al., 2012. Folliculin, the product of the Birt-Hogg-Dube tumor suppressor gene, interacts with the adherens junction protein p0071 to regulate cell-cell adhesion.. *PLOS One,* 7(11).

Menko, F. H. et al., 2009. Birt-Hogg-Dubé syndrome: diagnosis and mangement. *The Lancet Oncology,* Volume 10, pp. 1199-1206.

Meyer, P., Sergi, C. & Garbe, C., 2003. Polymorphisms of the BRAF gene predispose males to malignant melanoma. *Journal of Carcinogenesis,* 2(7), pp. 1-5.

Morell, R. et al., 1997. Apparent Digenic Inheritance of Waardenburg Syndrome Type 2 (WS2) and Autosomal Recessive Ocular Albinism (AROA). *Human Molecular Genetics,* 6(5), pp. 659-664.

Mota-Burgos, A. et al., 2013. Birt-Hogg-Dube Syndrome in a patient with melanoma and a novel mutation in the FLCN Gene. *International Journal of Dermatology,* 52(3), pp. 323-6.

Muller, F. L. et al., 2012. Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature ,* Volume 488, pp. 337-342.

Musgrove, E. A. et al., 2011. Cyclin D as a theraeutic target in cancer. *Nature Reviews Cancer,* Volume 11, pp. 558-572.

Nagy, A., Zoubarkov, D., Stupar, Z. & Kovacs, G., 2004. Lack of Mutation of the Follucilin Gene in Sporadic Chromophobe Renal Cell Carcinoma and Renal Oncocytoma. *International Journal of Cancer,* 109(3), pp. 472-475.

Nahorski, M. S. et al., 2010. Investigation of the Birt-Hogg-Dube tumour suppressor gene (FLCN) in familial and sporadic colorectal cancer.. *Journal of Medical Genetics,* 47(6), pp. 385-390.

Nahorski, M. S. et al., 2012. Folliculin interacts with p0071 (plakophillin-4) and deficiency is associated with disordered RhoA signalling, epithelial polarization and cytokinesis,. *Human Molecular Genetics,* 21(24), pp. 5268-5279.

Nakano, T. et al., 2003. Screening for mutations in BMP4 and FOXC1 genes in congenital anomalies of the kidney and urinary tract in humans.. *The Tokai Journal of experimental and clinical medicine,* 28(3), pp. 121-6.

Nan, H., Kraft, P., Hunter, D. J. & Han, J., 2009. Genetic variants in pigmentation genes, pigmentary phenotypes, and risks of skin cancer in Caucasians.. *International Journal of Cancer,* 125(4), pp. 909-917.

Nan, H. et al., 2011. Genome-wide association study identifies novel alleles associated with risk of cutaneous basal cell carcinoma and squamous cell carcinoma. *Human Molecular Genetics,* 20(18), pp. 3718-3724.

Narayanan, R. P. et al., 2012. IGFBP2 is a biomarker for predicting longitudinal deterioration in renal function in type 2 diabetes.. *Endocrine Connections,* Volume 1, pp. 95-102.

Ng, S. B. et al., 2010. Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genetics,* 42(1), pp. 30-35.

Ng, S. B. et al., 2009. Targeted capture and Massively Parallel Sequencing of Twelve Human Exomes. *Nature,* 461(7261), pp. 272-276.

Nickerson, M. L. et al., 2002. Mutations in a novel gene lead to kidney tumours, lung wall defects, and benign tumours of the hair follicle in patients with the Birt-Hogg-Dube syndrome.. *Cancer Cell,* 2(2), pp. 157-164.

Nucleics, 2013. *Exonuclease I - Shrimp Alkaline Phosphatase clean up of PCR products: Nucleics.* [Online]
Available at: http://www.nucleics.com/DNA_sequencing_support/exonucleaseI-SAP-PCR-protocol.html
[Accessed 1 July 2013].

Oetting, W. S. et al., 2009. The R402Q tyrosinase variant does not cause autosomal recessive ocular albinism. *American Journal of Medical Genetics,* 149(3), pp. 466-469.

Ozola, A. et al., 2013. Melanoma risk associated with MC1R gene variants in Latvia and the functional analysis of rare variants. *Cancer Genetics,* Volume 206, pp. 81-91.

Palmer, J. S. et al., 2000. Melanocortin-1 receptor polymorphisms and risk of melanoma: is the association explained soley by pigmentation phenotype?. *American Journal of Human Genetics,* 66(1), pp. 176-186.

Palmirotta, R. et al., 2010. Association Between Birt Hogg Dubé Syndrome and Cancer Predisposition. *Anticancer Research,* 30(3), pp. 751-757.

Pavlovich, C. P. & Schmidt, L. S., 2004. Searching for the hereditary causes of renal-cell carcinoma. *Nature Reviews Cancer,* Volume 4, pp. 381-393.

Perez Oliva, A. B. et al., 2009. Identification and Functional Analysis of Novel Variants of the Human Melanocortin 1 Receptor Found in Melanoma Patients.. *Human Mutation,* 30(5), pp. 811-822.

Pleasance, E. D. et al., 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature,* 463(7278), pp. 191-196.

Png, K. J., Halberg, N., Yoshida, M. & Tavazoie, S. F., 2011. A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells.. *Nature,* 481(7380), pp. 190-4.

Podsypanina, K. et al., 2008. Seeding and Progation of Untransformed Mouse Mammary Cells in the Lung. *Science,* 321(5897), pp. 1841-1844.

Prowse, A. H. et al., 1997. Somatic inactivation of the VHL gene in Von Hippel-Lindau disease tumours.. *American Jounral of Human Genetics,* 60(4), pp. 765-71.

Purdue, M. P. et al., 2011. Genoe-wide association study of rencal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nature Genetics,* 43(1), pp. 60-65.

Rana, B. K. et al., 1999. High polymorphisms at the human melanocortin 1 receptor locus. *Genetics,* 151(4), pp. 1547-1557.

Ray, K., Chaki, M. & Sengupta, M., 2007. Tyrosinase and ocular diseases: Some novel thoughts on the molecular basis of oculocutaneous albinism type 1. *Progress in Retinal and Eye Research,* 26(4), pp. 323-358.

Reiman, A. et al., 2012. Genetic Expression of Protein Array Studies of Folliculin-regulated Pathways.. *Anticancer Research,* 32(11), pp. 4663-4670.

Ricketts, C., Zeegers, M. P., Lubinski, J. & Maher, E. R., 2009. Analysis of Germline Variants in CDH1, IGFBP3, MMP1, MMP3, STK15 and VEGF in Familial and Sporadic Renal Cell Carcinoma. *PLoS ONE,* 4(6).

Robinson, S. J. & Healy, E., 2002. Human melanocortin 1 receptor (MC1R) gene variants alter melanoma cell growth and adhesion to extracellular matrix. *Oncogene,* Volume 21, pp. 8037-8046.

Ross, J., 1995. mRNA Stability in Mammalian Cells. *Microbiological Reviews,* 59(3), pp. 423-450.

Rozen, S. & Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. In: S. Krawetz & S. Misener, eds. *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Totowa(New Jersey): Humana Press Inc, pp. 365-386.

Saxonov, S., Berg, P. & Brutlag, D. L., 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Science,* 103(5), pp. 1412-1417.

Schödel, J. et al., 2012. Common genetic variants are the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nature Genetics,* 44(4).

Stetler-Stevenson, W. G., 1999. Matrix metalloproteinases in angiogenesis: a moving target for therapeutic intervention. *The Journal of Clinical Investigation,* 103(9), pp. 1237-1241.

Sulem, P. et al., 2007. Genetic determinants of hair, eyes and skin pigmentation in Europeans. *Nature Genetics,* Volume 39, pp. 1443-1452.

Su, L. et al., 2006. Genotypes and haplotypes of matrix metalloproteinase 1,3 and 12 genes and the risk of lung cancer. *Carcinogenesis,* 27(5), pp. 1024-1029.

Szumilas, M., 2010. Explaining Odds Ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry,* 19(3), pp. 227-9.

Tepass, U. et al., 2000. Cadherins in embryonic and neural morphogenesis. *Nature Reviews Molecular Cell Biology,* 1(2), pp. 91-100.

The Cancer Genome Atlas Research Network, 2013. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature,* 499(43), pp. 43-49.

Toro, J. R. et al., 1999. Birt-Hogg-Dube Syndrome A Novel Marker of Kidney Neoplasia. *Archives of Dermatology,* Volume 135, pp. 1195-1202.

Toro, J. R. et al., 2008. BHD mutations, clinical and molecular genetic investigations of Birt-Hogg-Dube syndrome: a new series of 50 families and a review of published reports.. *Journal of Medical Genetics,* 45(1), pp. 321-331.

Tringali, C. et al., 2012. The plasma membrane sialidase NEU3 regulates the malignancy of renal cell carcinoma by controlling β1 intergrin internalization and recycling. *The Journal of Biological Chemistry,* 287(51), pp. 42835-42845.

Truong, H. T., Dudding, T., Blanchard, C. L. & Elsea, S. H., 2010. Frameshift mutation hotspot identified in Smith-Magenis syndrome: case report and review of literature. *BMC Medical Genetics,* 11(142).

Turnbull, C. et al., 2010. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature Genetics,* 42(6), pp. 504-507.

UK, C. R., 2011. *Cancer Research UK About Cancer.* [Online]
Available at: http://www.cancerresearchuk.org/cancer-help/about-cancer/cancer-questions/how-many-different-types-of-cancer-are-there
[Accessed 4 March 2013].

Vahteristo, P. et al., 2010. No evidence for a genetic modifier for renal cell risk in HLCC syndrome.. *Familial Cancer,* Volume 9, pp. 245-251.

Valverde, P. et al., 1996. The Asp84Glu variant of the melanocortin 1 receptor (MC1R) is associated with melanoma.. *Human Molecular Genetics,* Volume 10, pp. 1663-1666.

van Steensel, M. A. M. et al., 2007. Novel Mutations in the BHD Gene and Absence of Loss of Heterozygosity in Fibrofolliculomas of Birt-Hogg-Dubé Patients. *Journal of Investigative Dermatology,* Volume 127, pp. 588-593.

Varela, I. et al., 2011. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature,* 469(7331), pp. 539-542.

Vernooij, M. et al., 2013. Birt–Hogg–Dubé syndrome and the skin. *Familial Cancer,* 10 January.

Vocke, C. D. et al., 2005. High Frequency of Somatic Frameshift BHD Gene Mutations in Birt-Hogg-Dubé-Associated Renal Tumours. *Journal of the National Cancer Institute,* 97(12), pp. 931-935.

Voisey, J., Box, N. F. & Van Daal, A., 2002. A Polymorphism Study of the Human Agouti Gene and its Association with MC1R. *Pigment Cell Research,* 14(4), pp. 264-267.

Voisey, J., Kelly, G. & Van Daal, A., 2003. Agouti Signal Protein Regulation in Human Melanoma Cells. *Pigment Cell Research,* 16(1), pp. 65-71.

Wang, L. et al., 2013. High level of FOXC1 expression is associated with poor prognosis in pancreatic ductal adenocarcinoma. *Tumour Biology,* 34(2), pp. 853-7.

Wang, R., Cukerman, E., Heng, H. H. & Liew, C. C., 1996. Identification of a locus of zinc finger genes in human chromosome 19q13.1-q13.3 region by fluorescence in situ hybridization.. *Somatic cell and molecular genetics,* 22(3), pp. 245-8.

Wang, Y. et al., 2004. Effect of the Codon 72 Polymorphism (c.215G>C, p.Arg72Pro) in Combination with Somatic Sequence Variant in the TP53 Gene on Survival in Patients with Advanced Ovarian Carcinoma. *Human Mutation,* Volume 24, pp. 21-34.

Wang, Z. et al., 2009. Regulation of immature dendritic cell migration by RhoA guanine nucleotide exchange factor Arhgef5. *Journal of biological chemistry,* 284(42), pp. 28599-606.

Warren, M. et al., 2004. Expression of Birt-Hogg-Dubé gene in mRNA in normal and neoplastic human tissues.. *Modern Pathology,* Volume 17, pp. 998-1011.

Wei, L. X. et al., 2013. High expression of FOXC1 is associated with poor clinical outcome in non-small cell lung cancer patients.. *Tumour Biology,* pp. 941-6.

Weiss, L., 1990. Metastatic Inefficiency. In: G. F. Vande Woude & G. Klein, eds. *Advances in Cancer Research.* San Diego(California): Academic press, pp. 159-161.

Wellcome Sanger Trust Institute, 2013. *COSMIC - Catalogue of Somatic Mutations in Cancer.* [Online]
Available at: http://cancer.sanger.ac.uk/cosmic/mutation/overview?id=1488330
[Accessed 15th September 2013].

Wellcome Trust Sanger Institute, 2013. *Wellcome Trust Sanger Institute.* [Online]
Available at: http://www.sanger.ac.uk/research/projects/cancergenome/
[Accessed 31 7 2013].

Wolf, E. et al., 2000. Effects of IGFBP-2 overexpression in vitro and in vivo. *Pediatric Nephrology,* 14(7), pp. 572-578.

Wood, A. M. & Stockley, R. A., 2006. The genetics of chronic obstructive pulmonary disease. *Respiratory Research,* 7(130).

Xia, G. et al., 2001. Regulation of vascular endothelial growth factor transcription by endothelial PAS domain protein 1 (EPAS1) and possible involvement of EPAS1 in the angiogenesis of renal cell carcinoma.. *Cancer,* 91(8), pp. 1429-1436.

Xia, L. et al., 2013. Overexpression of forkhead box C1 promotes tumour metastasis and indicates poor prognosis in hepatocellular carcinoma.. *Hepatology,* 57(2), pp. 610-24.

Xu, K. et al., 2012. A Comparative Study of Gene-Expression Data of Basal Cell Carcinoma and Melanoma Reveals New Insights about the Two Cancers. *PLoS One,* 7(1).

Xu, Z. Y. et al., 2012. FOXC1 contributes to microcasular invasion in primary hepatocellular carcinoma via regulating epithelial-mesenchymal transition.. *International Journal of biological sciences,* 8(8), pp. 1130-41.

Yokoyama, S. et al., 2011. A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. *Nature,* Volume 480, pp. 99-103.

Yoshida, T. et al., 2009. Association of genetic variants with chronic kidney disease in Japanese individuals.. *Clinical Journal of the American Society of Nephrology,* 4(5), pp. 883-890.

Zatyka, M. et al., 2002. Identification of cyclin D1 and other novel targets for the von Hippel-Lindau tumour suppressor gene by expression array analysis and investigation of cyclin D1 genotype as a modifier in von Hippel-Lindau disease.. *Cancer Research,* 62(13), pp. 3803-11.

Zhou, M. & He, H., 2013. Pathology of Renal Cell Carcinoma. *Current Clinical Urology,* pp. 23-41.

Zimmer, M., Doucette, D., Siddiqui, N. & Iliopoulos, O., 2004. Inhibition of Hypoxia-Inducible Factor is Sufficient for Growth Supression of VHL -/- Tumors. *Molecular Cancer Research,* Volume 2, pp. 89-95.