Genetics of an Ovine PKD Model

Bron Lett

A thesis submitted for the degree of

Master of Science

At the University of Otago, Dunedin.

New Zealand
ABSTRACT:

Autosomal recessive polycystic kidney disease (ARPKD) is an important cause of renal and liver related perinatal and neonatal death. ARPKD is presented as a gross enlargement of the kidneys due to the presence of multiple fluid filled cysts. When renal cysts present in the absence of additional abnormalities, ARPKD is often attributed to mutations in the PKHD1 gene which encodes the protein fibrocystin.

However many syndromes also present with renal cyst disease as one of a number of clinical characteristics, including Joubert syndrome, Bardet-Biedl syndrome, and Meckel-Gruber syndrome. These syndromes are all classed as ciliopathies which is a term describing diseases with structural or functional defects in the cell’s primary cilia, a sensory organelle that coordinates a large number of cellular signalling pathways.

We have identified an ovine model of polycystic kidney disease and have set up specific breeding and determined the inheritance of this disease to be recessive. Ovine SNP50 BeadChips, developed by Illumina in conjunction with the International Sheep Genomics Consortia (www.sheephapmap.org), were hybridized with DNA from affected sheep. Three candidate regions were identified by homozygosity mapping, one on each of chromosomes 4, 9, and 11.

In the region of concordant homozygosity on chromosome 9 we located the gene TMEM67 which is known to play a role in many ciliopathies. Upon sequencing of this gene two sequence variants were identified I638N and I644S. Bioinformatic analysis of both of these polymorphisms was unable to rule them out as the causative mutations in these sheep.

This study aimed to characterise the PKD sheep both phenotypically and genetically. The approach of using the SNP50 bead chip to identify regions of interest was crucial to the discovery of the sequence variants. At this time all the evidence supports the hypothesis that I638N and I644S are causative.
# TABLE OF CONTENTS

## CHAPTER 1: INTRODUCTION ................................................................. 6

1.1 HISTORY OF CILIA RESEARCH ....................................................... 6
1.2 STRUCTURE OF CILIA ................................................................. 8
1.3 CILIA MECHANICS ................................................................. 12
1.4 PRIMARY CILIA AND THEIR FUNCTION ........................................ 13
    1.4.1 PRIMARY CILIA IN DEVELOPMENT ....................................... 14
1.5 CILIOPATHIES ................................................................. 15
1.6 POLYCYSTIC KIDNEY DISEASE ................................................... 15
    1.6.1 ARPKD ................................................................. 16
    1.6.2 THE PCK RAT ........................................................... 19
1.7 CILIOPATHIC SYNDROMES ...................................................... 21
    1.7.1 PRIMARY CILIARY DYSKINESIA ......................................... 21
    1.7.2 PETINITIS PIGMENTOSA ................................................ 22
    1.7.3 BARDET-BIEDL SYNDROME .............................................. 23
    1.7.4 MECKEL-GRUBER SYNDROME ........................................... 23
    1.7.5 JOUBERT SYNDROME .................................................. 24
    1.7.6 NEPHRONOPHTHISIS ................................................... 24
1.8 GENETIC OVERLAP AND HETEROGENEITY .................................... 25
1.9 PUTATIVE FUNCTIONS OF TMEM67/MKS3 ..................................... 26
    1.9.1 TMEM67 IN ERAD ...................................................... 27
1.10 ANIMAL MODELS OF TMEM67 DEFECTS ...................................... 28
    1.10.1 THE WPK RAT ........................................................... 28
    1.10.2 THE BPCK MOUSE ..................................................... 29
1.11 CONGENITAL POLYCYSTIC KIDNEY DISEASE IN LAMBS ................. 30
CHAPTER 2: MATERIALS AND METHODS.................................32
  2.1 MATERIALS...........................................................................32
  2.2 ANIMALS................................................................................32
  2.3 MATINGS..................................................................................33
  2.4 TISSUE COLLECTION...............................................................34
  2.5 DNA EXTRACTION.................................................................35
  2.6 AGAROSE GEL ELECTROPHORESIS.................................36
  2.7 GENOTYPING (SNP50 BEADCHIP)....................................37
  2.8 HOMOZYGOSITY MAPPING.................................................37
  2.9 SHEEP GENOME BROWSER...............................................38
  2.10 PCR......................................................................................38
    2.10.1 PRIMERS........................................................................38
    2.10.2 PCR OPTIMISATION.......................................................39
  2.11 SEQUENCING........................................................................42
  2.12 KIDNEY HISTOLOGY............................................................42
  2.13 MUTATION ANALYSIS SOFTWARE.................................43

CHAPTER 3: RESULTS....................................................................45
  3.1 INTRODUCTION......................................................................45
  3.2 ANIMALS...............................................................................46
  3.3 DNA ISOLATIONS....................................................................47
  3.4 SNP50 BEADCHIP GENOTYPING....................................48
  3.5 HOMOZYGOSITY MAPPING...............................................50
  3.6 GENE IDENTIFICATION.........................................................51
Chapter 1 Introduction:

1.1 History of cilia research

The mammalian cilium has captured the attention and imagination of scientists for over a century, with K.W. Zimmerman first describing this organelle in 1898 (Wheatley, 2005). Despite being studied for so long, very little detailed knowledge about cilia could be obtained without advancements in both microscopy and molecular biology. This is why it took until 1954 for an in-depth study of the structure of cilia to be performed by Fawcett and Porter; this paper gave a clear depiction of motile cilia cytoskeletal structure and of their functions in the movement of fluid and debris along certain tissues (Satir & Christensen, 2008). Two years later De Robertis et al. found a form of non-motile cilia. These had a different ultraskeletal structure to motile cilia in that they lacked two central microtubules (De Robertis, 1956); these are now known as primary cilia. De Robertis also proposed that, given their lack of motility and because they were found in mammalian photoreceptors, they were most likely sensory organelles (De Robertis, 1956). Barnes (1961) also suggested that primary cilia played a sensory role due to their location on several sensory cell types. Until the 1960’s it was assumed that primary cilia were restricted to a small variety of mammalian tissues, however the use of electron microscopy soon changed this. Primary cilia were identified on adenohypophyseal cells by Barnes (1961) and on adrenocortical cells, pancreatic cells, testicular cells, kidney, and spleen cells by a large number of other researchers soon after (Currie and Wheatley 1966). A notable discovery was that primary cilia are found on neurons (Dahl, 1963), and as Karlsson showed in his 1966 paper, this was not a rare occurrence and in fact most, if not all neurons possess primary cilia (Karlsson, 1967). These discoveries soon caused many others to search a wide variety of cells for primary cilia, the end result being that primary cilia are nearly ubiquitous in mammalian tissues (Huang et al., 2006). To date there is no cell type that is known to possess both forms of cilia and, in 1962, Sergei Sorokin documented the two paths that cilia can take in their development to become either motile or primary cilia. He proposed that the development of cilia could be divided into three phases (Sorokin, 1962). The first phase involves a vesicle attaching to one end of a centriole; this vesicle is then invaginated by the growing ciliary bud turning the vesicle into the ciliary sheath. During the second phase many additional vesicles appear close to the centriole and cause the elongation of the ciliary bud into a shaft. The final phase involves the fusion of the ciliary sheath with the cell membrane. Sorokin hypothesised that
the stage at which a cilia become motile or primary is just before the start of the third phase (Sorokin, 1962). He based this on the observation that just before motile cilia emerge from the cell, granular vesicles are seen at the base of the cilia and it is thought that these play a role in the formation of the structures which allow for motility. The stimulus that causes the formation of the primary or motile cilia remains unknown (Wheatley, 2005).

Before the late 1970’s all information about primary cilia had been obtained by observation rather than experimentation. This changed in 1977 with the work of Albrecht-Buehler and colleagues. They studied the migration of ciliated cells across a surface that they had coated with gold particles. From the pattern in which the gold particles were removed it suggested that cells made 40 degree turns. 40 degrees happens to be the angular spacing between the microtubule doubles of the centriole (Albrecht-Buehler, 1977). Despite this paper, experimentation on the primary cilia was still lacking due to the insufficiency of the tools at the time. However valuable information could still be obtained by observation as shown with the 1979 paper by Reider et al. Before this paper there was only circumstantial evidence that the primary cilium was lost in dividing cells but no solid proof (Wheatley, 2005). By using PtK1 cells, which are kidney cells widely used for chromosome studies, it was shown clearly that primary cilia begin to be disassembled during the early phases of spindle formation during mitosis and are lost in a very short time span between prometaphase and metaphase (Rieder, Jensen, & Jensen, 1979). The re-growth of the primary cilia occurs in an equally prompt fashion, sometimes as early as late telophase (Wheatley, 2005). This finding gave a putative role for primary cilia in the control of cell division and cell cycle (Rieder et al., 1979). The exact role that primary cilia play in cell cycle regulation is still not known, however it has been demonstrated that cell cycle progression is linked to cilia shortening (Pan & Snell, 2007). This may however simply be due to the fact that the primary cilia needs to be disassembled before the centriole can perform its function in mitotic spindle formation and thus have little to do with the primary cilia controlling cell cycle progression signals (Pan & Snell, 2007).

After 1998 a revolution in cilia research began, this was due in part to primary cilia being tied to the disease situs inversus by Nonaka et al (1998). By knocking out the microtubule-dependant motor, KIF3B, Nonaka caused several disease phenotypes in mice. These
included morbidity before midgestation, growth retardation, and neural tube disorganisation (Nonaka et al., 1998). The most important phenotype was that the embryonic node (a part of the early embryo concerned with gastrulation) lacked primary cilia. Importantly these primary cilia were found to generate a leftward flow of extracellular fluid in wild type mice. This lead Nonaka et al. to propose that nodal cilia play a role in left-right patterning and that it was the loss of these cilia that resulted in randomisation of left-right symmetry, also known as situs inversus (Nonaka et al., 1998). In recent years it has been confirmed that the ability for the nodal cilia to produce a right-left flow is able to cause the activation of genes on the left side only, thus setting up left-right patterning (Buceta et al., 2005).

1.2 Structure of cilia

The function of cilia will be discussed shortly, but before we can have an understanding of cilium function it is important to look at the structures that make up cilia. Both primary cilia and motile cilia share a common internal structure known as the axoneme (Chodhari, Mitchison, & Meeks, 2004). The axoneme can be viewed as the backbone of cilia, as it is responsible for giving cilia their shape and much of the cilia structure and function is built around it. The axonemes of motile cilia consist of a scaffolding of microtubules in a nine plus two doublet (9x2 + 2) formation covered by the cellular membrane, that is to say that there are 9 pairs of microtubules that form a cylinder and in the middle of this cylinder are two more microtubules (Chodhari et al., 2004). The outer microtubules, which are composed of a complete A-tubule and an incomplete B-tubule, are connected by protein spokes to a sheath that surrounds the two medial microtubules, as well as being connected to each other by protein formations known as nexin links (Dillon & Fauci, 2000). These outer microtubule pairs are also linked to adjacent microtubule pairs by the protein complex dynein (Figure 1.0) (Dillon & Fauci, 2000). Dynein is a motor protein that plays a key role in the movement of motile cilia, the mechanism of how this occurs will be discussed in section 1.3.
Figure 1.0. The motile cilia. An illustrated example of the motile cilia including the structure of the axoneme and the centriole as well as other associated proteins.

As was stated earlier, primary cilia are distinguished structurally by their lack of the two central microtubules found in motile cilia, in addition most also lack the dynein motor complex (figure 1.1.) (Pazour & Witman, 2003). But as is often the case in biology there is an exception. Nodal cilia, mentioned earlier in their relation to situs inversus, are classified as primary cilia due to their lack of the two central microtubules. However, they have retained the dynein motor complex allowing for a partial “swirling” motion in place of the beating motion of motile cilia (Pazour & Witman, 2003). It turns out this swirling motion is crucial to their function in determining left-right patterning by setting up the right-left flow required to trigger the left side only gene cascade (Pazour & Witman, 2003).

At the base of the axoneme is a structure called the basal body, this is also known as the centrosome and the kinetosome. This structure was mentioned earlier in its role as the microtubule organising centre for cilia where it acts as the nucleation site for growing cilia (Moser, Fritzler, Ou, & Rattner, 2009). In addition to this it continues to play a role once the cilia has matured and is fully functional. For mature cilia the basal body acts as an anchor to the cell, and is set adjacent to the bottom of the axoneme (Chodhari et al., 2004). The basal body is derived from the centriole in a process that is yet to be identified, like the centriole the basal body is formed of microtubules in a series of nine triplets forming a hollow cylinder (9x3+0) (figure 1.0.) (Moser et al., 2009)
Figure 1.1. Electron micrographs demonstrating the lack of dynein arms and central microtubules in the primary cilium in comparison to motile cilium. (L) Transmission electron micrograph of a motile cilia showing the nine outer doublets (A), the dynein arms (B) and the two central microtubules (C). (R) Transmission electron micrograph of a primary cilium showing the nine microtubule doublets (D).

Ciliogenesis is dependent on a process known as intraflagellar transport or IFT. IFT, as the name suggests, acts to move materials into and out of the cilium. The work horses of this process are two separate motor proteins; kinesin-2 for anterograde (toward the tip of the cilia) transport and cDynein-1b for retrograde (toward the base of the cilia) transport (Figure 1.2.) (Cole et al, 2009). The structures that bind materials to the IFT motors consist of about 19 proteins. These proteins are highly conserved in all cilia hinting at their importance in ciliogenesis and normal cilium function (Cole et al., 2009). These coupling proteins form two separate complexes, referred to as ‘A’ and ‘B’. Complex A is made up of the proteins IFT144, 140, 139, and 122. Proteins IFT172, 88, 81, 80, 74/72, 57/55, 52, 46, 27, and 20 are associated with complex B. The nomenclature for these proteins comes from their molecular weight (Cole et al., 2009). Each of these proteins would appear to have an important function as knocking them out results in disease (Pazour & Rosenbaum, 2002). However the role that each protein plays has only just begun to be elucidated.

In ciliogenesis the IFT proteins are important as they are used to bring axoneme materials to the site of assembly at the tip of the cilia, thus without IFT cilia are unable to grow
They also play a role in removing used materials from the tip and returning them to the base of the cilia (Pazour & Rosenbaum, 2002).

Figure 1.2. Illustration of the intraflagellar transport system. This illustration shows the movement of the IFT system in relation to the basal body (retrograde) and the tips of the microtubules (anterograde).

All the proteins and structures thus far mentioned in relation to cilia are encased within the cell membrane, which subsequently projects into the extracellular space. The cilia membrane, like much of the cellular membrane, has many receptors and transmembrane proteins associated with it (Pazour & Rosenbaum, 2002). The type and density of receptors is variable depending on the tissue the cilia is found in and the subsequent role of the cell it is located on. Two such examples are the SST3 somatostatin receptor, and the 5-HT6 serotonin receptor. The SST3 somatostatin receptor is a seven transmembrane G-protein-coupled receptor that regulates several cell processes. In particular it is important in regulating the release of growth hormone in certain areas of the brain, thus controlling the proliferation of neurons. It also plays a role in neuronal excitability as it regulates the release of the neurotransmitters GABA and acetylcholine (Pazour & Rosenbaum, 2002).

The SST3 receptor, unlike other members of its family, is localised to the primary cilia of
several areas of the brain including the amygdala, hippocampus, thalamus, and cerebral cortex (Pazour & Rosenbaum, 2002). The 5HT6 serotonin receptor is important in the serotonin pathway of the brain which has influence in many neurological processors such as mood, memory, and learning. Currently it is being looked at as a possible pharmacological target to treat schizophrenia, anxiety and obesity. 5HT6 is strongly localised to the primary cilia of the islands of Calleja but also in the nucleus accumbens, olfactory tubercle, and striatum. Having these receptors localized to the primary cilium means that they are able to be projected into small extracellular spaces and be activated by a local release of their ligand, allowing them to relay a message back to the cell body which may then be able to propagate a signal to neighbouring cells (Pazour & Rosenbaum, 2002).

Looking at the structure of cilia demonstrates how complex an organelle it is and that many genes are required to function correctly for cilia to be effective in their jobs. The obvious consequence of this is that there are a large number of components that can fail resulting in disease.

1.3 Cilia mechanics

Having discussed the structure of cilia, the way in which these structures come together to create the characteristic movement of motile cilia will now be examined.

The axoneme creates movement through the action of dynein arms that are bound to the A-tubule and connect to an adjacent microtubule. Dynein is a motor protein that generates sliding of the microtubules in an ATP-induced unidirectional power stroke resulting in bending of the cilia. When the dynein motors are activated in a coordinated fashion it can result in the characteristic beating of the flagella or motile cilia, however the exact mechanism of how dynein activation is controlled in such a manner remains unknown.

Although structurally cilia and eukaryotic flagella are the same it has been suggested that the difference between flagella and motile cilia is the way in which they move, with flagella moving in a propeller like motion and motile cilia moving in a back and forth, wave like motion. Flagella are used for locomotion of single free cells such as sperm cells; this is in contrast to the function of motile cilia in movement of liquids relative to a tissue, such as the movement of mucus in the trachea (Dillon & Fauci, 2000).
1.4 Primary cilia and their function

Primary cilia are present in almost every cell in a vertebrate’s body usually as just one per cell (Pazour & Witman, 2003). Despite being so widespread the exact function of primary cilia is not well known or understood, and many hypotheses have been formed to try and explain them. For a long time it was thought that primary cilia were simply vestigial organelles, however this does not fit well with them being so widespread in the body, and so highly conserved across so many species (Pazour & Witman, 2003). It was also hypothesised that they played a role in cell cycle regulation, given that they are removed just before spindle formation and are present on the cell at all other times. However this could also be due to the fact that they rely on the centrosome, which is also required to play a role in spindle formation. In order for chromosomal migration to occur the primary cilia must be disassembled (Pan & Snell, 2007). The most well supported current hypothesis is that primary cilia are sensory organelles. This hypothesis has gained widespread acceptance in recent years. It was originally proposed because lower eukaryotes use cilia as sensory organelles in their visual and olfactory systems (Pazour & Witman, 2003). Since then more evidence about the role of the primary cilia as a sensory organelle has come to light from studies of various disease models caused by cilia defects. One piece of experimental evidence that supports primary cilia having a function comes from the Tg737orpk mouse. This model develops polycystic kidney disease (PKD), and using scanning electron microscopy it was revealed that the primary cilia of these mice were unable to form properly resulting in PKD (Pazour & Witman, 2003). Given the hypothesis that primary cilia are sensory organelles, it was expected that there would be many receptors and receptor associated channels that are localized to them. Confirmation of this is found in examining the autosomal dominant polycystic kidney disease genes PKD1 and PKD2. These genes encode the proteins polycystin-1 and polycystin-2 respectively. Polycystin 1 is a receptor protein that contains a large extracellular N-terminal region, multiple transmembrane domains, and an intracellular C-terminal tail. Polycystin-2 is a cation channel belonging to the TRP family. These proteins act together to form a receptor-channel complex that plays a role in intracellular calcium homeostasis (Gallagher, Germino, & Somlo, 2010). The polycystin protein complex also plays a role in cell-cell/matrix interactions and renal tubular development. Its role in renal tubule development is a possible explanation for why defects in this gene result in PKD. The
polycystin protein complex has been localized to the primary cilia (Gallagher et al., 2010) allowing it to project into and sense the extracellular space.

1.4.1 Primary cilia in development

Given its role in cell signalling and in sensing the environment, it is little wonder that the primary cilium is important in mammalian embryo development. An interesting example of this is the role cilia play in left-right patterning during early mammalian embryo development (Nonaka et al., 1998). The cilia that are important for this are known as the nodal cilia because they reside in a small triangular indentation on the embryo known as the node. To call these primary cilia is not entirely correct. As although they have the 9+0 axoneme of primary cilia they also have dynein arms and display some motility which is not characteristic of primary cilia (Nonaka et al., 1998). The motility would however seem to be crucial to the proper patterning of the embryo as demonstrated by the iv mutant mouse in which the nodal cilia lack motility. This is due to a mutation in the gene that encodes the left-right dynein and subsequently the mice develop situs inversus (Nonaka, Shiratori, Saijoh, & Hamada, 2002). Situs inversus is a condition in which the organs are on the opposite side of the body to what they should be; proper left-right asymmetry in mammals is caused by a cascade that is trigged by a right to left flow pattern caused by the nodal cilia (Klysik, 2008). Although it is clear that the nodal flow is important, the exact reason is not known. It is thought that it activates a gene cascade in only the left side cilia; this may be by mechanical displacement or by the movement of morphogen containing vesicles (Klysik, 2008). Evidence for the case of mechanical displacement came in an experiment in which the flow over the node could be manipulated (Nonaka et al., 2002). When the normal leftward flow of wild type mice was reversed they developed situs inversus, and when iv mice had a leftward flow applied they developed normally (Nonaka et al., 2002). These experimental conditions are unlikely to have allowed for a morphogen gradient to be set up, thus it is more likely that mechanical activation of the primary cilia at the edge of the node cause a gene activation cascade (Klysik, 2008). This hypothesis is also lent weight by the ability of primary cilia to cause Ca2+ influx in result to fluid flow, and that when this Ca2+ influx is blocked, by mutating either the PKD1 or PKD2 genes which encode the PC1-PC2 channel complex, it can result in situs inversus (Cantiello, 2003). The role of the primary cilia in situs inversus was one of the key factors in causing the huge increase of interest in primary cilia research since 1998 when Nonaka et al first implicated that primary cilia played a role in Kartenger’s syndrome (Wheatley, 2005).
1.5. Ciliopathies

In addition to situs inversus and PKD, there are several other developmental conditions that primary cilia are implicated in. Examples are hepatic defects such as cirrhotic livers; brain defects such as occipital encephalocele and agenesis of the corpus callosum. In addition to cystic kidneys, primary cilia defects also cause cytogenesis in several other tissues such as the liver, the epididymis, the pancreas, and the bile ducts. Collectively these diseases are known as ciliopathies (Klysik, 2008). Recently primary cilia have also been tied to obesity and some forms of cancer. Given that these are not developmental diseases, it is apparent that the role of the primary cilia continues right through an individual’s life (Nigg & Raff, 2009).

From the wide gamut of conditions which primary cilia are associated with we see that its function is integral to proper development and maintenance of the mammalian body. However because of the complex nature of the primary cilia and the vast number of proteins required for it to operate there are many mutations that can cause ciliopathies. Interestingly, mutations in different genes can cause different cross-sections of symptoms despite the fact that they are all having an effect on the same organelle. These range from the PKD1, PKD2, PKHD1 and PKHD2 genes which cause polycystic kidney disease as the main symptom with a small number of extra renal symptoms (Ibraghimov-Beskrovnaya & Bukanov, 2008), to the rarer genes such as TG737, MKS3 and the NPHP genes that are implicated in many syndromes and cause a wide variety of defects usually in conjunction with PKD (Davenport et al., 2007; Klysik, 2008).

1.6. Polycystic kidney disease

One of the most important primary cilia related diseases is PKD as this group of diseases are the most commonly inherited, lethal disorder, affecting 12.5 million people worldwide. PKD is a congenital disease that affects the tubules of the kidney, impairing their function and causing cysts to form (Wilson. et al, 2004). Cystogenesis occurs when cilia are damaged and lose their function. The cells they are attached to undergo a change in cell morphology and signalling. In PKD the alteration of kidney cells results in the formation of cysts (Rosenblum. et al, 2007). In many cases of PKD these cysts result in end stage renal disease which requires a kidney transplant or dialysis in order for a patient to live (Wilson. et al, 2004). PKD comes in two major forms; autosomal recessive PKD (ARPKD) and autosomal dominant PKD (ADPKD). ADPKD is the most common form affecting
about one in four hundred live births. It is a late onset disease, usually only occurring in an individual’s late 40’s. PKD is characterised by a gross enlargement of the kidneys due to multiple fluid filled cysts, and usually results in end stage renal failure requiring either continued dialysis or a kidney transplant. Most cases of ADPKD can be attributed to the PKD1 gene that encodes polycystin; the other 20% are associated with the PKD2 gene (Wilson. et al, 2004). The polycystin protein has a large extracellular domain as well as several transmembrane domains; it likely plays a role in cell-cell/matrix interactions and may play a role in intracellular calcium homeostasis. In development it is important in renal tubule development, hence why defects in this gene cause ADPKD.

1.6.1 Autosomal Recessive Polycystic Kidney Disease

ARPKD is much less common owing to it being recessively inherited. It affects about 1 in 20,000 individuals, and is characterised by pre-or neonatal lethality (Harris, Torres, David, Md, & Martin, 2008). Like ADPKD, ARPKD has enlarged kidneys due to multiple cysts. However, in ARPKD, cyst formation predominately occurs during foetal development. Because of this, many neonates die of pulmonary hypoplasia due to oligohydraminos. ARPKD is most commonly due to defects in the PKHD1 gene, but can also be caused by mutations in the PKHD2 gene or in several other loci such as TG737 (Harris et al., 2008). In addition to the two main types of PKD there are a number of other syndromes that have ARPKD as one of many clinical characteristics, including Joubert, Bardet-Biedl, and Meckel-Gruber syndrome (Quinlan, Tobin, Beales, & Robert, 2008). These syndromes will be discussed in greater detail in section 1.6.

ARPKD is characterized by enlargement of the kidneys caused by dilation of the collecting ducts and congenital hepatic fibrosis (Harris. et al, 2004 ). Respiratory failure is the major cause of neonatal death with 30% of affected neonates dying due to the enlarged kidneys causing pulmonary hypoplasia (Harris. et al, 2004 ). However advances in medicine are making it possible for some cases of pulmonary hypoplasia caused by ARPKD to survive. Sumfest et al, 1993. describes aggressive surgery to treat three infants with ARPKD and of these, two managed to survive. Neonates that do survive often develop hypertension and 20-45% develop end stage renal failure by the age of 15 (Harris. et al, 2004 ).

In humans the gene responsible for the majority of ARPKD cases has been identified as PKHD1. This is a large gene covering a region of 472kb and containing a minimum of 86 exons and is primarily expressed in the kidney (Onuchic. et al, 2002). The protein it
encodes is 4074aa long, 447kDa and is known as fibrocystin. This protein is encoded by the longest open reading frame which encompasses 67 exons (Xiong. et al, 2002). As yet the function of fibrocystin is unknown. It has a large extra cellular domain with a signal peptide, a single transmembrane domain and a short intracellular domain (Xiong. et al, 2002). The extracellular domain of fibrocystin contains twelve, immunoglobulin like folds which are found in transcription factors and plexins (Xiong. et al, 2002) and in several receptor proteins. Its cytoplasmic tail also contains several possible phosphorylation sites (Onuchic. et al, 2002). The structure of fibrocystin, along with the proteins it shares homology with, suggests that it plays a role as a receptor, possibly involved in modulating the terminal differentiation of collecting ducts (Onuchic. et al, 2002).

The PKHD1 gene was initially mapped to chromosome 6p21-cen by Zerres et al. in 1995 and later refined by Mucher et al to chromosome 6p21.1-p12. They identified the gene transcript was 11.6kb long and had 61 exons with the initiation codon lying in exon 2(Xiong. et al, 2002). This group named the gene product of PKHD1 “tigmin” and predicted that it was 3396 amino acids long. The expression patterns of PKHD1 were also examined by Xiong et al. using multiple tissue northern blots they found that there was strong expression in the kidney (which they also found to be localized to the nephron), weak expression in the liver, and little expression in the pancreas and foetal brain. They also noted two hybridisation bands which hinted at alternative transcripts (Xiong. et al, 2002).

At the same time as Xiong et al. identified the PKHD1 gene two other groups also identified the gene, Ward et al 2003. and Onuchic et al 2002. The Ward paper entitled “The gene mutated in Autosomal recessive polycystic kidney disease encodes a large receptor like protein” found an almost identical sequence, apart from 15bp at 5’-UTR (Ward. et al, 2002). The gene also had an identical expression pattern. Ward et al. named the protein product fibrocystin, and Onuchic et al. named it polyductin. Onuchic et al. also showed that there were at least 86 exons contained in the PKHD1 gene and that through alternative splicing the main product, polyductin, was produced from the longest reading frame.

Xiong et al. focused on the cDNA product of PKHD1, whereas Onuchic focused on the genomic DNA, thus the differences in the number of exons each group found (67 and 86 respectively). However Onuchic started looking at transcripts while they found the 67 exon transcript described by Xiong et al. they also went beyond this and found several other
transcripts made by alternative splicing of the PKHD1 gene (Onuchic. et al, 2002). However they state that the actual number of alternative splice variants is undoubtedly much higher as they did not perform an exhaustive analysis to detect for splice variants (Onuchic. et al, 2002). Onuchic et al. then went further and performed a mutation analysis on 25 individuals affected by ARPKD as well as 60 healthy control subjects. For this they focused on the transcript of the longest open reading frame of 67 exons. They found potentially pathogenic variants in 42% of the chromosomes screened (defined as a variant present in patients but not in control individuals). They regarded this as strong evidence that they had indeed found PKHD1.

They also found homology to several proteins including hepatocyte growth-factor receptor (HGRF) and several plexins (Onuchic. et al, 2002). Additionally they identified the ITP immunoglobulin folds; however they were unable to definitively comment on their function. But they did note that the ITP domains are shared by two families of protein, DNA transcription factors and single-pass cell-surface receptors that are members of the Sema super-family of proteins (i.e., HGFR, Ron, and the large family of plexins) (Onuchic. et al, 2002). The arrangement of the ITP domains more closely resembles that of the receptor proteins such as HGFR, indicating that the function of polyductin may be related. However they also pointed out that polyductin is missing other domains that are present across all other members of the Sema receptor super-family indicating that it has a novel function (Onuchic. et al, 2002). Interestingly they also located PbH1 repeats; these repeats are most commonly associated with plant bacterial virulence allowing the breakdown of the cell wall polysaccharides (Onuchic. et al, 2002). An arginine-glycine-aspartate (RGD) domain was found which indicates possible function in cell adhesion; additionally three putative cAMP/cGMP phosphorylation sites were identified within the cytoplasmic carboxyl terminus (Onuchic. et al, 2002). All these differing motifs presented a protein that could have any number of functions from cell adhesion, to carbohydrate recognition. This information combined with the possibility of many splice variants meant that PKHD1 could be performing many different roles and may been acting as a receptor, a ligand, or a membrane associated protein.
1.6.2 The PCK Rat

Although many different rodent models of ARPKD have been made none were found to have been caused by a region orthologous to the ARPKD interval in humans (Lager. et al, 2001). In 2001 the PCK rat model was described as being phenotypically similar to human ADPKD with collecting duct derived cysts and hepatic cyst disease (Lager. et al, 2001). Upon examination of the rat it was found that its phenotype more closely resembled ARPKD. Using linkage markers Ward et al. mapped the causative gene to chromosome 9.

To test whether this region was the ortholog of the human ARPKD interval, they identified polymorphisms in the rat ptd011 gene (An ortholog of the human PTD011 gene which had been mapped to the ARPKD interval) (Ward. et al, 2002). When affected animals were typed for polymorphisms in this gene they found complete correspondence between homozygotes and the PCK phenotype, thus showing that this region is the ortholog to the human ARPKD interval (Ward. et al, 2002). Comparing the expression of the human and mouse PKHD1 gene we see that the gene is expressed in a similar pattern being highly expressed in the kidney with small amounts present in the liver and pancreas and no expression in the brain and colon (Ward. et al, 2002).

Comprehensive mutational studies have been carried out on the PKHD1 locus (Rossetti. et al, 2003) and thus far 119 mutations have been identified (see table 1 below for details of the mutations). The rates of detection per allele were relatively low ranging from 47% to 61%. This low rate has been attributed to the variability of diagnosis and severity of ARPKD, as there was a higher detection rate in the more severe cases and rates as low as 40% in moderate cases (Rossetti. et al, 2003). It was also suggested that there is genetic heterogeneity, but this was deemed unlikely because of the overwhelming evidence linking ARPKD to PKHD1. One interesting explanation is that the mutations causing the milder phenotypes are present in unscreened exons that are part of splice variants for the protein.
Table 1. 119 mutations identified in the PKHD1 gene that have been published (Rossetti. et al. 2003)

Of the 119 mutations that have been described; 70 are missense, 27 are frame shifting insertions or deletions, 12 are nonsense mutations, 9 are splicing changes, and one inframe deletion has been found (Rossetti. et al. 2003). Of the mutations 36% were found to occur just once. The mutation T36M was found in all the studies and accounted for 16.9% of all the mutated alleles identified (Rossetti. et al. 2003). The alleles R496X and 9689delA were found at a higher than expected rate in Finnish and Spanish populations respectively.
The most common types of mutation causing ARPKD are truncating and missense mutations. Of the 89 patients where both mutations identified these two types account for the mutations in 80 of the patients (table 2) (Rossetti. et al, 2003).

<table>
<thead>
<tr>
<th>Types of mutation</th>
<th>Number of patients affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truncating and missense</td>
<td>35</td>
</tr>
<tr>
<td>Two missense</td>
<td>26</td>
</tr>
<tr>
<td>Two truncating</td>
<td>19</td>
</tr>
<tr>
<td>Splicing and truncating</td>
<td>5</td>
</tr>
<tr>
<td>Splicing and missense</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2. Types of mutations present in the 89 patients with both mutant alleles identified, and the number of patients affected by those mutations.

Because of complex combinations of heterozygote mutation it is difficult to determine genotype/phenotype relationships. One that was identified is that two truncation mutations resulted in severe ARPKD, with all 19 patients, in these studies, who had two truncating mutations dying perinatally (Rossetti. et al, 2003). Genetics based diagnostics of ARPKD is made difficult because of the large gene size, variable mutation detection rates, alternative splicing, and allelic heterogeneity (Rossetti. et al, 2003).

1.7. Ciliopathic syndromes.

Given that primary cilia are found on many of the body’s cells it is understandable that when there is a defect it causes a wide range of abnormalities. This is particularly well demonstrated by the number of multi-symptom syndromes that are associated with primary cilia defects. These syndromes often have overlapping phenotypes, which is to be expected given that they are all associated with defects in the same organelle (Klysik, 2008). As stated earlier, cilia are nearly ubiquitous throughout the body. This is reflected in the wide array of tissues that ciliopathies affect, including: epithelium of the respiratory tract, oviduct, testes, brain, kidney, eye, inner ear, and olfactory epithelium (Klysik, 2008).

1.7.1 Primary Ciliary Dyskinesia

Primary ciliary dyskinesia (PCD) is a recessively inherited disorder that, despite its name, affects the motile cilia of many epithelial tissues. The defects are prominent in the respiratory tract, the sinuses, several parts of the ear, and fallopian tubes (Chodhari et al.,
All of these tissues require proper motile cilia function and disruptions in this can lead to many problems such as respiratory infection due to poor mucus secretion (Chodhari et al., 2004). One interesting form of PCD is Kartagener syndrome, which accounts for an estimated 50% of PCD cases (Klysik, 2008). First described in 1933 by its name sake, Manes Kartagener, it is characterised by a triad of symptoms. These are chronic sinusitis, bronchiectasis, and situs inversus. The first two symptoms are commonly shared in all forms of PCD as they are caused by the reduced movement of the motile cilia in the sinuses and bronchial tree respectively (Klysik, 2008). This results in a build up of secretions, which in the sinuses causes infection and in the bronchial tree causes respiratory problems. The third symptom, situs inversus, was described earlier as a reversed orientation of organs on the left-right axis. Looking at the symptoms described for PCD it is not surprising that genes that have been linked with the disease are found to be important for cilia motility and maintaining structure. The two genes that have been shown to have an association with PCD are DNAI1 and DNAH5 or dynein intermediate chain 1 and dynein heavy chain 5 respectively. Both these genes, as is evident from their names, encode proteins associated with the dynein arm which, as was discussed previously, is a protein complex crucial in creating the movement of motile cilia (Klysik, 2008).

### 1.7.2. Retinitis Pigmentosa

One of the more peculiar places in which cilia defects occur is the eye. The most well known form of ciliopathy of the eye is retinitis pigmentosa (RP) which is a degenerative process of the retina that primarily affects rod cells (Pagon, 1988). Because rod cells are used for peripheral and night vision, loss of these is a major characteristic of RP. RP is the most common form of inherited blindness and there are currently thirty-two genes known to be associated with the disease (Wang et al., 2005). About 60% of RP cases still have no known genetic cause, and as there are over 70 loci that cause photoreceptor degeneration in Drosophila it is likely that there are still many more genes yet to be found in humans. Because of the large number of genes there is also a large amount of heterogeneity and RP can occur as either an isolated disease or as part of a syndrome (Wang et al., 2005). In healthy individuals cilia play an important role in the transport of pigments in photoreceptors. This is because rod cells are separated into inner and outer segments. The inner segment is where pigments required for photoreception are made, and the outer segment is where they perform their role (De Robertis, 1956). Connecting the two segments is a cilium. Thus many of the genes that are associated with RP are cilia related.
intraflagellar transport proteins such as *Kif3A* (Klysik, 2008). Mutations in this gene cause the build up of rodopsin and arrestin in the pigment producing inner segment of rod cells. It is hypothesised that this build up of pigment causes apoptosis, which on a large scale causes the degeneration of the retina (Klysik, 2008).

### 1.7.3. Bardet-Biedl Syndrome

One of the syndromes that sometimes has RP as a symptom is Bardet-Biedl syndrome (BBS). The pleiotropic nature of Bardet-Biedl syndrome is an excellent example of how a single gene mutation affecting the primary cilia can cause multiple system problems (Beales & Luciano, 2004). The main clinical characteristics of Bardet-Biedl syndrome are obesity, retinitis pigmentosa, polydactyly, mental retardation, hypogonadism, and renal dysfunction. BBS also has a large number of secondary symptoms that can present such as: diabetes mellitus, heart disease, hepatic fibrosis, ataxia, hypertonia, and cardiovascular anomalies (Beales & Luciano, 2004). However these are much rarer. As well as its phenotypic heterogeneity, BBS is also highly genetically heterogeneous with twelve genes thus far being identified. These genes are known as BBS1-BBS12 (Klysik, 2008). Adding to its genetic complexity is the identification of triallelic inheritance in some families. This means that in some cases, in order for BBS to present, three mutations in the BBS genes need to be inherited (Beales & Luciano, 2004). Several genes, such as MKS3, have also been found to be modifiers of the disease with different alleles causing variation of the rarer phenotypes (Leitch et al., 2008). The disease phenotype of BBS at birth is seen to strongly resemble another ciliopathic syndrome, Meckel-Gruber syndrome (Karmous-Benailly et al., 2005). In particular both syndromes are found to share polycystic kidneys, polydactyly, and liver defects. The cross over between the two is so great that a 2005 paper by Karmous-Benailly *et al.* found that many foetuses diagnosed as having Meckel-Gruber syndrome in fact had mutations in one of the BBS genes suggesting that the antenatal presentation of BBS may mimic Meckel-Gruber syndrome.

### 1.7.4. Meckel-Gruber syndrome

Meckel-Gruber syndrome (MKS) is an autosomal recessive disease that, as stated above, is often characterised in humans by polydactyly, bilateral renal cyst dysplasia, and liver defects (Velasco, 2009). Although it is often characterised by these symptoms the classical triad associated with human MKS has occipital encephalocele, polydactyly, and bilateral renal cyst dysplasia (Velasco, 2009). Like other ciliopathies there is a certain amount of
phenotypic heterogeneity and a large amount of overlap with other conditions. These conditions include, but are not limited to, ARPKD, trisomy 13, Smith-Lemli-Opitz syndrome, hydroolethralus syndrome, Senior-Loken syndrome, Bardet-Biedl syndrome, and oral-facial-digital syndrome type 1 (Otto et al., 2009). Many of these are also ciliopathies. Most of the confirmed cases of MKS have been linked to three loci in humans, MKS1, MKS2, and MKS3 which have been mapped to 17q23, 11q13, and 8q21.13-q22.1 respectively. From the three loci two genes have been identified, MKS1, and TMEM67 in the MKS3 loci (Velasco, 2009). TMEM67 encodes a protein known as Meckelin, while the MKS1 protein retains the name of its gene. Both of these proteins have been implicated in the formation of cilia and seem to play a role in the actions of the basal body and its migration. However the exact details of the functions of these proteins are unknown (Otto et al., 2009).

1.7.5. Joubert Syndrome

The MKS3 loci/TMEM67 gene that is responsible for MKS type 3 has also been implicated in another syndrome. Joubert syndrome (JS) is another recessively inherited disorder caused by disruptions to the function of cilia. JS is characterised by hypoplasia of the cerebella vermis, developmental delays, hypotonia, ataxia, abnormal breathing patterns, and abnormal eye movement (Ferland et al., 2009). In some cases of JS a combination of occipital encephalocele and postaxial polydactyly are reported. These two symptoms together are commonly associated with MKS and are included in the classical triad of the syndrome. In addition to TMEM67, the genes AHI1 and CEP20 have also been identified as causative in JS (Ferland et al., 2009).

1.7.6. Nephronophthisis

NPHP is a ciliopathy that also presents with cystic kidneys. It is the most common cause of end-stage renal disease (ESRD) in the first three decades of life (Hildebrandt et al., 2009). The three clinical forms are defined by the median age of ESRD onset, being 1 year for infantile, 13 years for juvenile, and 15 years for adolescent (Hildebrandt et al., 2009). The phenotypic difference between NPHP and PKD is the area of the kidney in which cysts arise. In PKD cysts are found evenly spread throughout the entire kidney, while the cysts of NPHP form at the corticomedullary junction (Hildebrandt et al., 2009). Like PKD, NPHP can also be found with extra-renal symptoms. When coupled with retinitis pigmentosa it is known as Senior-Loken syndrome. An association with Joubert syndrome
is also identified when NPHP is coupled with cerebellar vermis aplasia. Other symptoms that can occur with NPHP include liver fibrosis and ocular apraxia type cogan. These extra-renal symptoms are found in about 10% of all NPHP patients (Hildebrandt et al., 2009). NPHP is a highly genetically heterogeneous disorder with 9 genes so far being associated (Otto et al., 2009). These genes are named NPHP1 through to NPHP9. The functions of these genes are yet to be fully elucidated.

1.8. Genetic Overlap and Heterogeneity

A 2009 paper by Otto et al. entitled “Hypomorphic mutations in Meckelin (MKS3/TMEM67) cause nephronophthisis with liver fibrosis (NPHP11)” came to the conclusion that MKS, JS, and nephronophthisis (NPHP) are allelic disorders of the same gene, TMEM67 (Otto et al., 2009). In this paper they looked to identify genes associated with the subset of NPHP patients with liver involvement. To achieve this they used whole genome scanning on 3 affected individuals who are part of a consanguineous family. Using a 50k SNP microarray and linkage mapping they obtained a significant LOD score in the area containing the gene TMEM67. In the 3 affected individuals they identified a homozygous missense mutation in this gene. Four more novel mutations were identified in TMEM67 when a world-wide cohort of NPHP was examined. When looking at causative MKS mutations in TMEM67 they found that they were predominantly nonsense or truncating mutations. And when 120 patients with JS were sequenced for the TMEM67 gene, 7 different mutations were identified. Of the mutations that caused the NPHP and liver fibrosis they found that these were usually missense mutations, leading to the conclusion that missense mutations in TMEM67 cause a milder phenotype than truncating mutations (Otto et al., 2009).

A summary of the genes associated with Bardet-Biedl syndrome (BBS), Meckel-Gruber syndrome (MKS), Joubert syndrome (JS), and Nephronophthisis (NPHP) is included below. As shown in the Venn diagram below all these syndromes are highly genetically heterogeneous and many of the causative genes are shared among the diseases.
Figure 1.4. A Venn diagram showing the genetic heterogeneity and the overlap between five ciliopathic syndromes; Bardet-Biedl syndrome (BBS), Meckel-Gruber syndrome (MKS), Joubert syndrome (JS), Nephronophthisis (NPHP) and Senior-Loken syndrome (SLS).

1.9. Putative Functions of TMEM67/MKS3

Of interest in figure 1.4. is the area of overlap over the MKS3 gene. This area includes the four syndromes that are associated with PKD: MKS, BBS, JS, and NPHP. The gene MKS3 is also known as TMEM67 and encodes the protein Meckelin. The gene itself is 47kb long, contains 28 exons and is greatly conserved across all higher organisms. Meckelin is 995 amino acids long and is predicted to have 3-7 transmembrane domains with a large extracellular N-terminus containing a signal peptide and cystine rich domain. The intracellular C-terminus contains a coiled-coil domain. Crystallography to determine the structure of this protein is difficult due to it being associated with the cell membrane. The function of Meckelin is yet to be fully elucidated; however it is known to play a role in the migration of the centriole which is an important step in cilia formation. Dawes et al (2006) hypothesised that meckelin might be a non-canonical wnt-receptor that plays a role in actin or tubulin cytoskeletal rearrangements. From their observations that meckelin localises to the plasma membrane and MKS1 localises to both the centrosome and the plasma membrane Dawes et al., (2006) suggested that MKS1 might provide the link between meckelin at the membrane and p160ROCK, which is a protein involved in actin stress fibre
formation, at the centrosome thus assisting in its migration. This would fit well with the known role of meckelin in basal body migration. Williams et al., (2010) found that ciliogenesis did not require TMEM67, however when this gene was disrupted the cilia was elongated and chemoreception was abnormal. The cilia phenotype that Williams et al., (2010) found in *C. elegans* matches that reported for MKS3 mutations in human, mice, and rats. In addition to the elongation of the cilia in mutant *C. elegans*, Williams et al., (2010) also noted the presence of multiple centrosomes and multiple cilia. They also found that TMEM67 and MKS1 interact with nphp-1 and nphp-4 to influence proper positioning, orientation, and formation of cilia. Nphp-1 and nphp-4 have both been implicated in the ciliopathies nephronophthisis and Senior-Loken syndrome. The authors putatively identified TMEM67 as acting upstream of nphp-4 as when both were knocked out the effect was not greater than when TMEM67 was knocked out by itself. Williams et al 2010 also came to the conclusion that the two pathways share similar functions allowing for some redundancy, and that the MKS1/MKS3 pathway has a partial role in the biology of the cilium membrane as opposed to the axoneme.

1.9.1 TMEM67 in ERAD

An additional role for TMEM67 was found by Wang et al., (2009). This paper was looking at the mechanisms of the endoplasmic reticulum-associated degradation (ERAD) pathway which is involved in removal and degradation of misfolding proteins. When they stressed the ERAD pathway either genetically by expressing misfolding proteins or pharmacologically they found that expression of TMEM67 was increased 5 fold. Subsequent knockouts of TMEM67 in lung cells resulted in the build up of misfolding proteins that would normally be removed by the ERAD pathway. This build up eventually led to formation of cytotoxic aggregates and activation of apoptosis. The authors found that TMEM67 interacts with p97, which is a translation initiation factor involved in cell cycle regulation, to aid in the removal of misfolded proteins from the ER to the cytosol. They concluded that the evidence supports a model in which TMEM67 links the ER luminal quality control machinery with the cytosolic degradation apparatus. This function would appear to have very little to do with ciliogenesis, however these findings may give a clue as to another role of TMEM67 in cilia. Given that on the ER Meckelin plays a role in trafficking of proteins, it may have a similar role in the cilia. This would be plausible with the observations of Williams et al 2010 that mutations in TMEM67 didn’t stop ciliogenesis and instead caused cilia to elongate. If TMEM67 is regulating the control of proteins into
the forming cilia and it becomes defective it is conceivable that an over abundance of cilia proteins could locate to the cilia tips and become incorporated causing larger than intended cilia.

Summarily, TMEM67 plays a role in both ciliogenesis and in the endoplasmic reticulum-associated degradation pathway. It has a putative function in protein trafficking in ERAD, and could possibly play a similar role in ciliogenesis. This is in addition to its possible role of tethering transition fibres to the membrane to allow for basal body migration.

1.10. Animal Models of TMEM67 defects

As mentioned earlier the PCK rat is a disease model for ARPKD caused by the pkhd1 gene. However with so many genes and syndromes that have ARPKD as a symptom it is useful to have several disease models so that a full understanding of each disease and how the diseases relate to each other can be obtained. To date there are two model organisms for defects in the TMEM67 gene, the wpk rat and the bpck mouse.

1.10.1 The wpk Rat

The wpk rat is the most well known of these organisms. It was first described in 2004 by Gattone et al as a model organism for polycystic kidney disease. It was described as developing proximal tubule and collecting duct cysts in utero, these cysts were seen to eventually dominate the kidney. Several extrarenal characteristics were also noted. These included hypoplasia of the thymus and spleen and CNS malformations such as agenesis of the corpus callosum and hydrocephalus. The authors mapped the wpk gene to an 11cM interval on the rat chromosome 5 close to a marker known as D5rat73. Within this region was a known PKD modifier locus. In 2006 Smith et al refined this region to a 7.5 cM locus between the markers D5Rat131 and D5Rat210. They achieved this by analysis of 566 F2 and other informative backcrossed animals resulting from the crossing of Wistar wpk rats and wild-type Brown Norway rats. To further refine the area new markers in the 2.5 Mb area surrounding the D5Rat73 locus were identified. The marker SWE5P, which is telomeric to D5Rat73, was found to be recombinant in one mouse. And on the centromeric side the SL11.1 marker was found to be recombinant. This gave a region of 0.6 cM or a physical region of 2.05Mb. 13 conserved genes were identified in this region, and upon sequencing of the novel gene LOC313067 they found a substitution, P394L, in exon 12. From this they identified LOC313067 as the wpk gene. In noticing the phenotypic
similarities between the wpk rat and MKS patients the authors sequenced the human ortholog of wpk, TMEM67, in several MKS families. Upon identifying correctly segregating mutations in these families they came to the conclusion that TMEM67 was the causative gene in MKS3 individuals and in the wpk rat.

1.10.2. The bpck Mouse

The bpck mouse was initially described by Cook et al 2009. They described a mouse model with polycystic kidney disease, hydrocephalus, and morbidity by 3 weeks. They identified the causative mutation to be a spontaneous 295-kb deletion that includes the TMEM67 gene.

Although the mouse model does have some characteristics in common with MKS3 it is also lacking some that are very common in human patients. These are hepatic involvement, polydactyly, occipital encephalocele, and cerebellar defects. However it is seen to have elongated cilia which are found in both the humans and the wpk rat.

The deletion was mapped to a 0.6cM interval on chromosome 4 between the markers D4Mit261 and D4Mit19. This deletion encompassed six candidate genes. These were: cadherin 17 (Cdh17), protein phosphatase 2C magnesium dependent catalytic subunit (Ppm2c), an unidentified gene predicted from adult testis cDNA, Riken cDNA 1700123M08, Tmem67, and Riken cDNA C430048L16.

Looking at the candidate genes, TMEM67 stood out to the authors given the phenotypic similarities between the bpck mouse and MKS3 patients. In order to test the hypothesis that TMEM67 was the causative gene they performed transgenic rescue by using four overlapping BAC’s covering the deleted region. Inheriting the BAC’s containing TMEM67 rescued the mice and allowed for survival past a year with no signs of PKD, hydrocephalus, or elongated cilia.

All the other genes contained in the BAC’s with TMEM67 were also in BAC’s that did not rescue the phenotype thus ruling them out as being causative. This led to the conclusion that although multiple genes were knocked out by the deletion the only one causing the phenotype of interest was TMEM67.
1.11. Congenital Polycystic Kidney Disease in Lambs

A 2005 paper by AC Johnstone et al titled “Congenital polycystic kidney disease in lambs” describes two New Zealand flocks of sheep that exhibited a high incidence of polycystic kidney disease causing prenatal death. PKD in New Zealand lambs had been recorded since 1950 (Johnstone, Davidson, Roe, Eccles, & Jolly, 2005). However adequate descriptions of the pathology and epidemiology were not available. In order to study this disease two south island flocks were identified that had a high prevalence of affected lambs. Flock A, which is a Perendale x Perendale-East Friesian crossbred flock located in Canterbury, had 15 lambs present with PKD in 2002. DNA profiling was performed using blood from the adult lambs and renal tissue of three affected lambs. This allowed for the creation of an experimental group using confirmed carrier animals. Two rams that were identified as sires of PKD lambs and 71 of their daughters, including three dams of the PKD lambs were selected for mating in 2003. The next year the two Rams and 25 of the ewes were transported to Massey University where they could be mated in more controlled conditions. Eight of the ewes were presumed to be obligate heterozygotes due to them having previously had PKD lambs; about half of the remainder of the ewes were expected to be carriers due to being daughters of the carrier rams. These matings produced 15 affected lambs in 2003 and 10 affected lambs in 2004.

Flock B, a Coopworth flock located in Southland, had 20 PKD lambs in 1999 and 50 affected lambs the following year (Johnstone et al., 2005). Steps to eliminate the phenotype from the flock were taken including culling of all but five rams and the purchase of an additional eight unrelated rams. Despite this 20 affected lambs were born in 2001 and 12 more in 2002. The parentages of three affected lambs were assigned using DNA from renal tissue and the blood of putative dams and sires. One ram was identified as being the sire of all three affected lambs and two ewes were confirmed as carrier dams. In 2000 and 2001 planned matings between the known carriers was performed, the carrier ram was also mated with 25 additional unrelated ewes. From the planned matings two affected lambs were born from the known carrier ewes. No affected lambs were born from the unrelated ewes.

Extensive assessment of the phenotype of PKD lambs was performed on 10 affected lambs born in 2004 from flock A. The lambs were noted as having a distended abdomen in which enlarged kidneys were palpable. Dissection of the abdomen revealed greatly enlarged
kidneys. Multiple fluid filled cysts were found to affect the majority of the tissue upon sectioning of the kidney with very little distinction between the cortex and medulla being apparent. The urinary bladder of the affected lambs was noted to be a narrow cylindrical structure; this malformation was thought to be due to a lack of urine distending to the bladder during gestation. Hepatic defects such as portal fibrosis and fibrous tissues spreading beneath the capsule were noted, however the extent of fibrosis varied. Pancreatic involvement was noted in all affected lambs, with the normal lobular structure of the pancreas being replaced by a network of dilated ductal structures that were directly connected to the extrahepatic bile duct which was also highly cystic and dilated. Additionally the epididymis of the lambs underwent cystic changes; however there were no other changes in the sex organs of either sex. Histopathology showed that the normal architecture of the kidney was disrupted by cystic tubules. The cysts formed at all levels of the nephron and caused the surrounding interstitum to become reduced. In the liver there was a great variation in the severity of the lesions that affected the parenchyma. The least-developed cysts tended to be around the portal zone, and the greatest cystic changes where observed in the areas where intra and extra hepatic ducts converge.

Renal cysts in conjunction with cystic tissue in other organs are features of many forms of both animal and human PKD and from the numbers of lambs born with PKD the authors drew the conclusion that it was a recessively inherited disorder.

Since the publishing of the Johnstone paper additional work has been carried out on the PKD flock established from flock B at AgResearch, Mosgiel. Work by summer student, Samara Bretherton, examined the cilia of the affected lambs and found them to be stunted, however recent work by Tony Poole has identified elongated cilia and it is thought that the stunted cilia Samara noticed were broken. In 2005 PhD candidate Jane Wilson-Wheeler undertook sequencing of putative causative genes in the PKD flock. From the phenotype it was hypothesised that PKHD1 and TG737 were likely candidates. However sequencing of these genes didn’t return any mutations.

The aim of this thesis was to identify the causative gene and mutation in the PKD flock. This is a flock of sheep that produces animals affected by bilateral renal cyst dysplasia with hepatic involvement. A thorough description of the phenotype associated with mutant animals would also need to be established in order to compare these lambs to human and other animal models of this disease, as well as the phenotype of flock A.
Chapter 2

Materials and Methods

2.1. Materials

A full list of the chemicals and kits used for this project can be found in appendix A with the formulations of the solutions used in appendix B

2.2. Animals

Several different breeds of sheep from multiple flocks were used for this research. The flock that was used for the majority of the breeding consists of New Zealand Coopworth sheep that were derived from a South Island flock shown to have a high prevalence of autosomal recessive polycystic kidney disease (ARPKD). This flock, referred to as flock B in Johnstone et al., (2005), produced about 100 PKD lambs over the span of 4 years, despite efforts to reduce the prevalence such as culling of known carriers and buying in new rams for mating. A new flock was set up for the purpose of studying the form of recessive PKD that these sheep exhibit, it was derived from flock B which was located on Winton farm in the South Island of New Zealand. This new flock, which will be referred to in this thesis as the PKD flock, was kept in a private field at the AgResearch Invermay campus in Mosgiel. In order to establish the PKD flock, confirmed carriers from flock B had to be selected for mating. It was rare for the birth of affected lambs to be witnessed. As such carrier ewes were usually identified by their close proximity to the lambs that had died of the disease. In order to identify carrier rams, DNA finger printing of all the rams, affected lambs, and suspected ewes present in flock B was performed by sub-contractors under the direction of M. Eccles in 2000. This identified one carrier ram, 96-642, and confirmed the status of the ewes. It also identified another putative carrier ram, 01-4. These carriers were then moved to Invermay in 2001. The use of these sheep for research purposes is covered by the University of Otago animal ethics approval number 88/07.

In addition to the animals from flock B, samples were taken from animals derived from flock A which, as described in Johnstone et al., (2005), were Perendale sheep from a mid-Canterbury farm that had produced 15 PKD lambs in 2002. In 2004, 25 ewes and 2 rams from flock A were moved to Massey University, Palmerston North. Eight of the ewes were
presumed to be obligate carriers due to having given birth to affected lambs. Approximately half of the remaining ewes were thought to be carriers as they were the daughters of rams 5004 and 5005; the sires responsible for the affected lambs in previous matings, making them obligate carriers themselves.

A large number of unrelated wild-type ovine DNA samples were also provided by Agresearch Invermay. These samples were collected from five different breeds of sheep: Romney, Coopworth, Perendale, Texel, and Cheviot. The sheep that these samples were collected from are owned and cared for by AgResearch.

2.3. Matings

Planned matings were set up from 2005 onwards. The PKD flock normally consists of between one to three carrier rams that are mated with a large number of carrier ewes and potential carrier ewes. Only recently has it been possible to genotype live animals. As such, the status of many animals was unknown when breedings were set up. This meant that a large number of matings were unsuccessful in producing affected offspring as only the status of obligate carriers was known.

When the PKD flock was first moved to Invermay in 2001, carrier ram 96-642 was mated with the 4 carrier ewes as well as 25 additional ewes already present at Invermay. This produced one affected PKD lamb; however 96-642 had to be euthanized due to back problems and sperm collected were not viable for in vitro fertilization. Between 2002 and 2004 further breeding was carried out. However no PKD lambs were born.

In 2005 eight additional carrier ewes were purchased from flock B and added to the PKD flock. Ram 98-8587, which was a potential carrier from a Winton farm, and two sons of 96-642 were mated with 15 known carrier ewes and 9 daughters of 96-642. This round of breeding produced one PKD lamb sired by 98-8587, which confirmed its status as a carrier.

2006 saw 98-8587 being crossed with all available ewes, 11 of which were confirmed carriers. This resulted in 4 PKD lambs. In 2007, three sons of 98-8587 (6005, 6026, and 6093) were mated with 15 ewes resulting in 3 PKD lambs, one from each Ram, making them all obligate carriers. All three carrier rams were again mated in 2008, this time with 22 ewes. However, only 2 PKD lambs were born in 2008, both from 6005. The 2009
breeding consisted of 6005 mating with 24 ewes, 6 of which were known carriers, and resulted in 5 PKD lambs.

The breeding information is summarized in table 1

<table>
<thead>
<tr>
<th>Year</th>
<th>no. rams</th>
<th>no. ewes (known carriers)</th>
<th>no. lambs</th>
<th>no. PKD</th>
<th>Sires</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>1</td>
<td>29 (4)</td>
<td>40</td>
<td>1</td>
<td>96-642*</td>
</tr>
<tr>
<td>2004</td>
<td>1</td>
<td>16 (7)</td>
<td>30</td>
<td>0</td>
<td>01-4,</td>
</tr>
<tr>
<td>2005</td>
<td>3</td>
<td>24 (15)</td>
<td>47</td>
<td>1</td>
<td>01-2, 04-27, 98-8587*</td>
</tr>
<tr>
<td>2006</td>
<td>1</td>
<td>12 (8)</td>
<td>35</td>
<td>4</td>
<td>98-8587*</td>
</tr>
<tr>
<td>2007</td>
<td>3</td>
<td>15 (7)</td>
<td>23</td>
<td>3</td>
<td>6005*, 6026*, 6093*</td>
</tr>
<tr>
<td>2008</td>
<td>3</td>
<td>22 (9)</td>
<td>40</td>
<td>2</td>
<td>6005*, 6026, 6093</td>
</tr>
<tr>
<td>2009</td>
<td>1</td>
<td>24 (6)</td>
<td>36</td>
<td>5</td>
<td>6005*</td>
</tr>
</tbody>
</table>

* had PKD offspring

2.4. Tissue collection

Prior to 2009 all tissue from the PKD flock was collected post-mortem as lambs were found dead in the field up to 12 hours following birth. This made the tissue unacceptable for many forms of processing including cell culture and histology. Until 2009 it was assumed that all affected lambs died during birth as none had been found alive until the lambing of that year. In 2009 better protocols, as outlined below, were put in place to allow for tissue to be collected as soon as possible after the lambs had died, giving better quality samples and making extensive processing permissible.

In the years preceding 2009 the status of both the lambs and many of the ewes was unknown; this meant that there was no way of knowing if a lamb was going to be affected until many hours after it was born. An attempt was made to rectify this for the 2009 lambing season. Genotyping for the putative mutation was not available at the time of
mating therefore pregnant sheep had ultrasound analysis performed on them during the third trimester. These scans were able to predict with a high degree of certainty if any foetal lambs suffered from PKD. If a ewe from the PKD flock was confirmed to be carrying an affected lamb then it was moved into holding pens for careful observation. These ewes were then monitored several times a day for signs of labour by the farm manager.

When an affected lamb was born it was euthanized, if required, and then bought into the laboratory for dissection. The dissection was extensive and required a team of investigators as many tissues and organs were removed and sampled. The sheep was first cleaned and its ears were removed for DNA isolation. Next the abdominal cavity was opened, and in most affected cases, the kidneys dominated the abdomen and were removed first. The kidneys were cut in half on the transverse plane, and then smaller sections were made and put into either 4% paraformaldehyde or formalin in 10% NBF fixing solution. Tissue was also collected to develop epithelial and fibroblast primary cell lines. The liver was sampled in a similar fashion, followed by the pancreas, the bile ducts, and the epididymis. After the relevant tissues had been sampled from the abdominal cavity the knee joint was opened up to allow for cartilage to be taken, this was sampled in a similar fashion to the other tissues. Lastly the heads were removed whole and frozen for MRI scans to look for brain deformities. During this entire procedure photos were taken to document any abnormalities.

2.5. DNA extraction

Ovine tissue collected from the MKS flock prior to 2009 had been stored at -20 since its collection. DNA was able to be isolated from many of these tissues using a Promega Wizard DNA isolation kit (Cat #. A1120). This method involved the digestion of the tissue with 17.5 μl of proteinase K over night. The protein was then removed using the protein precipitation solution. Next the DNA was cleaned using isopropanol and precipitated using ethanol. The pellet was air dried and finally resuspended in DNA rehydration solution. The full method used can be found in the Promega Wizard DNA isolation kit manual. This method was used to isolate DNA for the SNP50 BeadChip, PCR, and sequencing.

DNA was ideally isolated from the ear or kidney of the sheep, but in some instances liver tissue was used. DNA was able to be extracted from ear tissue using a similar method to
that described for mouse tail DNA extractions in the Promega kit, which was described above. However kidney or liver tissue had to be physically broken up prior to DNA extraction, this was done by cutting the tissue into smaller sections with sterilized surgical scissors. To test if the isolation of DNA was successful, DNA was run on agarose gels and the amount and purity of DNA was assessed using a nano drop spectrophotometer. If the DNA came out as a smear instead of a tight, bright band it was considered unsuccessful as it was likely that the DNA was degraded thus making it inappropriate for use in PCR. If the 260/280 ratio was lower than 1.80 on the nano drop it was considered to be unacceptable for hybridisation to the SNP50 BeadChip (see section 2.7).

Blood samples were collected from the Massey flock and DNA was extracted from these by Pam Cornes using the Qiagen Qiaamp DNA blood midi kit (Qiagen, Venlo, Netherlands). This is a column spin kit that involves the lysis of whole blood, then the binding of the DNA to a stationary phase, several washes with buffer, and finally elution of the DNA. DNA from 30 additional, unrelated wild type sheep samples was extracted and provided by Fiona Sanggang of AgResearch Invermay.

2.6. Agarose gel electrophoresis

DNA was visualised using agarose gel electrophoresis. 1.5% agarose gels were cast and run using a Bio-Rad mini-sub cell gel tank (Bio-Rad Laboratories, Hercules, California, USA). Agarose gel was made by dissolving SeaKem LE agarose powder (Cambrex bioscience, Rockland Inc. Rockland, ME, USA) in 1 x TAE buffer containing 50 µl ethidium bromide (Invitrogen, Carlsbad, California, USA). Typically, 200ml of gel was made at a time. This involved adding 3 grams of agarose powder to a Schott bottle and then filling up to the 200ml mark with buffer. This mixture was then microwaved until the agarose powder had completely dissolved.

DNA samples (both genomic and PCR product) were mixed with ~1 µl of loading dye (0.2% Bromophenol blue, 0.2% Xylene Cyanote) and loaded into the wells. Fragments were sized using a 1kb plus DNA ladder at 2 ng per gel (Invitrogen, Carlsbad, California, USA). The gel underwent electrophoreses at 100V until adequate separation of the fragments was achieved; this was judged by the distance that the leading dye front, caused by bromophenol blue, had travelled in the gel. The gel was visualised using the UV light setting in a BIORAD Universal Hood and imaged using the Quantity One Gel Doc system.
2.7. Genotyping (SNP50 bead chip)

Fourteen affected animals and one carrier from the MKS flock were chosen to be sent away for genotyping using the Ovine SNP50 bead chip developed by AgResearch and Illumina (www.illumina.com). This chip has 54,241 evenly spaced probes that target single nucleotide polymorphisms (SNPs) found in the sheep genome. The chip uses Illumina’s Infinium HD assay. This assay uses 50-mer probes that selectively hybridise the loci of interest. The probe sequence stops one base short of the target SNP, the sample DNA is then used as a template for single base extension of the probe using a labelled nucleotide. The labelled nucleotides are then detected by the iScan system allowing for the allele to be called. Individually SNP’s are not very informative as they only have two alleles. Nevertheless when in a high density they are able to create informative haplotypes that allow for recessive traits to be identified with as few as 10 affected animals.

2.8. Homozygosity mapping

Genotyping data from the ovine SNP50 beadchip was analysed by using Excel spreadsheets that had been custom programmed by Dr. David Markie to identify regions of autozygosity shared by all affected animals. Excel was used to align the SNPs from all affected animals and then identify regions of concordant homozygosity that extend for a predetermined number of SNPs. This was achieved by instructing Excel to change the base calls into binary code, where a homozygous call equals 1 and a heterozygous call equals 0. With this, homozygous base calls were highlighted in red. Next SNPs that had concordant homozygosity across all animals were assigned a 1 and highlighted in red. Finally the percentage of concordant homozygosity as a function of the window size was established. This was achieved by moving a frame along the data for concordantly homozygous SNPs. If the window size was set to ten and the number of concordantly homozygous SNPs was eight then that region would be marked as 80%, then the frame would move along one SNP and mark that region of ten SNPs. If a region was 100% concordantly homozygous then it would be marked in yellow indicating a region of interest.
2.9. Sheep genome browser

Regions of concordant homozygosity were examined for putative genes using the sheep genome browser (Ovine genome assembly v1.0) from the Australian sheep gene mapping website (http://rubens.its.unimelb.edu.au/~jillm/jill.htm). This genome browser takes the virtual sheep genome version 1.0 and aligns it against the SNPs from the SNP50 ovine beadchip, the human reference sequence, and the bovine reference sequence. The alignment with the human and bovine sequence is useful for identifying putative genes.

2.10. PCR

2.10.1 Primers

Primers were designed using Primer3 v0.4.0 (Rozen and Skaletsky, 2000). Partial sequence for the gene was obtained from AgResearch, but given that the sheep genome was not fully sequenced at the time, large amounts of the sequence were unknown. It is also known that because the current virtual sheep genome was aligned using the bovine genome as a template there may be discrepancies that would not become obvious until PCR or resequencing was performed. Initially primers were designed to try and amplify as much of the putative gene as possible in an attempt to identify the mutation. In regions of the gene where the sequence was not complete options for primers were limited and as such some primers were made with relatively low melting temperatures and GC content.

However a paper was published later that showed that mutations in the MKS3 gene were either non-sense or missense. So the exons of the gene were identified by taking the sequence of the human exons from ensemble and BLASTing (NCBI blast, http://blast.ncbi.nlm.nih.gov/Blast.cgi) them against the sheep sequence, specifically the align programme was used. The exons that were found that were not already covered by working primers then had new primers made to amplify them. Melting temperature and GC content played a more important role in the design of these new primers with GC content always being 40% to 60%. If an exon was unable to be mapped to the sheep sequence then primers were made to sequence unknown regions that might contain the missing exons, these were identified using the Bovine and human sequence and looking for homologous intronic sequence.

Primers were ordered from the Invitrogen website using their custom primer service (http://www.invitrogen.com/site/us/en/home.html). Once the primers were received from
Invitrogen they were diluted to 100μM for a stock solution, 5μM for PCR, and 1 micromolar for sequencing.

2.10.2 PCR optimisation.

Once the primers had been designed and diluted they were optimised by altering the annealing temperature and the length of the extension step. A full list of primers can be found in table 3. The PCR conditions were as follows:

<table>
<thead>
<tr>
<th>Per reaction:</th>
<th>MQH_2O</th>
<th>15 μL</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X PCR Buffer</td>
<td>1.8 μL</td>
<td></td>
</tr>
<tr>
<td>50mM MgCl₂ (1.5mM final)</td>
<td>.60 μL</td>
<td></td>
</tr>
<tr>
<td>10mM dNTPs (0.2mM final)</td>
<td>.40 μL</td>
<td></td>
</tr>
<tr>
<td>Taq polymerase</td>
<td>.20 μL</td>
<td></td>
</tr>
<tr>
<td>Primer (5μM)</td>
<td>1 μL</td>
<td></td>
</tr>
<tr>
<td>Sample DNA</td>
<td>1 μL</td>
<td></td>
</tr>
<tr>
<td>Total PCR volume</td>
<td>20 μL</td>
<td></td>
</tr>
</tbody>
</table>

18 μL of reaction mix was added to 0.2 mL PCR tubes containing 1 μL of 5 μM primer and 1 μL of DNA to give a total PCR reaction volume of 20 μL.

**PCR Program:**
1. 95°C for 2 minutes.
2. 95°C for 30 seconds.
3. 45-60°C for 30 second.
4. 72°C for 1-4 minutes.
5. Repeat steps 2 to 4.
6. 72°C for 5 minutes
7. Pause at 5°C

Steps 2-4 were repeated 35 times. Step 3 is the annealing step of the cycle, the temperature for this step changes depending on optimal temperature for the primers used. Step 4 is the extension step, the time used for this step is dependent on the length of the fragment being amplified; longer fragments require more time to be replicated.
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
<th>TM</th>
<th>Primer name</th>
<th>Primer sequence</th>
<th>TM</th>
</tr>
</thead>
<tbody>
<tr>
<td>f1</td>
<td>AAAACTATCGGGGAAATGTCA</td>
<td>59</td>
<td>r1</td>
<td>CATGGAATCTGGGTCTGCA</td>
<td>59</td>
</tr>
<tr>
<td>f2</td>
<td>AAGGAAATCAAATGTGAATGTCAG</td>
<td>59</td>
<td>r2</td>
<td>TGGTTTGTGGATTTGCGCTTG</td>
<td>60</td>
</tr>
<tr>
<td>f3</td>
<td>CCACCTCCCATATTTCTCTTC</td>
<td>59</td>
<td>r3</td>
<td>AAGCCAGATGCCCCTTCTGTA</td>
<td>59</td>
</tr>
<tr>
<td>f4</td>
<td>AGTGGGGAAAAAGCACACAT</td>
<td>60</td>
<td>r4</td>
<td>TGGTCAGGGTGGTGAATCTG</td>
<td>60</td>
</tr>
<tr>
<td>f5</td>
<td>GAAATGGTAAACAGCATCATTACTG</td>
<td>60</td>
<td>r5</td>
<td>CTCCTTTGGGCATCCTTATA</td>
<td>60</td>
</tr>
<tr>
<td>f6</td>
<td>ATCCCACTACATCCACTTC</td>
<td>59</td>
<td>r6</td>
<td>TGAGAAACCAACACCCAAAAGA</td>
<td>60</td>
</tr>
<tr>
<td>f7</td>
<td>GGTTCCATGAATTTCACTCC</td>
<td>59</td>
<td>r7</td>
<td>AGCCACCTGCAACAGGAT</td>
<td>59</td>
</tr>
<tr>
<td>f8</td>
<td>CCAAAACCTCTTTGGCTACA</td>
<td>57</td>
<td>r8</td>
<td>TGAACCTTAGTGAACATTTAACCTC</td>
<td>57</td>
</tr>
<tr>
<td>f9</td>
<td>TGGATCATACTGATTTCTTTC</td>
<td>59</td>
<td>r9</td>
<td>CATTTTCATAAAACCTGCTGGA</td>
<td>59</td>
</tr>
<tr>
<td>f10</td>
<td>AAAGATCTGAACATATCATTGT</td>
<td>60</td>
<td>r10</td>
<td>TGGACCTTTTTGCGGAAGTA</td>
<td>60</td>
</tr>
<tr>
<td>f11</td>
<td>TGGAACACCACTTTACGTTTTT</td>
<td>58</td>
<td>r11</td>
<td>TATGAGTGGTCATCGTCTTC</td>
<td>59</td>
</tr>
<tr>
<td>f12</td>
<td>AAATTTCTTCGAACTGACTGGTAT</td>
<td>57</td>
<td>r12</td>
<td>CACTGACCTGGCTACCTTACCA</td>
<td>59</td>
</tr>
<tr>
<td>f13</td>
<td>TCCTGTAGAGCTAGGCTCTGTA</td>
<td>59</td>
<td>r13</td>
<td>ACTCAGCCTGGAATCCTGTA</td>
<td>59</td>
</tr>
<tr>
<td>f14</td>
<td>AGCAATGCTGCTGCAAAAAAC</td>
<td>60</td>
<td>r14</td>
<td>CGAGACCTAAACTGCGCATAA</td>
<td>59</td>
</tr>
<tr>
<td>f15</td>
<td>GGGCAATAAGTGATTTCTGCT</td>
<td>58</td>
<td>r15</td>
<td>TCTCAGGGTCAACTCTCA</td>
<td>60</td>
</tr>
<tr>
<td>f16</td>
<td>GGGCTGGGGAGTGAAATAATA</td>
<td>58</td>
<td>r16</td>
<td>CTAATTAAGAAGTAAACACAGCAGA</td>
<td>58</td>
</tr>
<tr>
<td>f17</td>
<td>AAATGAGCTAAACAGGTTTTTCATGC</td>
<td>58</td>
<td>r17</td>
<td>ATGAAAGCATTTATCTCTCTCTTTT</td>
<td>57</td>
</tr>
<tr>
<td>f18</td>
<td>CAATCTCTTCTGAAATGGAGAA</td>
<td>59</td>
<td>r18</td>
<td>TCTGGCTTTTTGTGACTCCA</td>
<td>59</td>
</tr>
<tr>
<td>f19</td>
<td>TCCCTCTAGAGGGGAAATGGG</td>
<td>60</td>
<td>r19</td>
<td>TCAATGGGTCGCTGGGAAG</td>
<td>60</td>
</tr>
<tr>
<td>f20</td>
<td>CTTGGCTACGCTGTGGTTGCTT</td>
<td>59</td>
<td>r20</td>
<td>TAGCTCTTACAGGCGGATT</td>
<td>60</td>
</tr>
<tr>
<td>f21</td>
<td>GTCAACACAGCTGACGATCAGTA</td>
<td>60</td>
<td>r21</td>
<td>CATGATACCTCTCAAAAAGCA</td>
<td>60</td>
</tr>
<tr>
<td>f22</td>
<td>ACAACACGAATCCTGAGACA</td>
<td>59</td>
<td>r22</td>
<td>TTTTGCCAAATGGAAGGA</td>
<td>59</td>
</tr>
<tr>
<td>f23</td>
<td>AATTCTATCAATAGTTTTTTTACCTCCA</td>
<td>57</td>
<td>r23</td>
<td>TACCTTTGTGATCAATAATGCA</td>
<td>59</td>
</tr>
<tr>
<td>f24</td>
<td>CGTCTCCCAATCGATAAAGAA</td>
<td>60</td>
<td>r24</td>
<td>ATTGGTCTCTGCTGCCCA</td>
<td>59</td>
</tr>
<tr>
<td>f25</td>
<td>TCAGTGGGAAAGACGCTCAGC</td>
<td>59</td>
<td>r25</td>
<td>TGGAGAAACTGGGTTGATATAAGT</td>
<td>58</td>
</tr>
<tr>
<td>f26</td>
<td>CCACACAAGGTTTTTTACTGGC</td>
<td>57</td>
<td>r26</td>
<td>TGGAACTCGACAGAATGAGG</td>
<td>59</td>
</tr>
<tr>
<td>f27</td>
<td>GCCTCATAAAATGATAATTCTCTGTA</td>
<td>59</td>
<td>r27</td>
<td>TGTCACAATCTGAGTAGGTTTTTCTATA</td>
<td>58</td>
</tr>
<tr>
<td>f28</td>
<td>TGATCTTGACATAGGGGAGGA</td>
<td>60</td>
<td>r28</td>
<td>TTTTCTGGCCAAATTTGTC</td>
<td>60</td>
</tr>
<tr>
<td>f29</td>
<td>GGACATAGGCGCACAATTAA</td>
<td>60</td>
<td>r29</td>
<td>CATATGTCGTATCACAACCTTTGGCA</td>
<td>59</td>
</tr>
<tr>
<td>f30</td>
<td>TTAATTTTCTGATTATGGACCACCTG</td>
<td>57</td>
<td>r30</td>
<td>TTTGAGATACATGGTGCTGC</td>
<td>59</td>
</tr>
<tr>
<td>f31</td>
<td>GAACTGGACCAGCTTGGAGG</td>
<td>59</td>
<td>r31</td>
<td>GGAGAAATCCAGGAGAATG</td>
<td>59</td>
</tr>
<tr>
<td>f32</td>
<td>ACCCCAAAATGACCTTCCCA</td>
<td>58</td>
<td>r32</td>
<td>CAGCGGCTAGTGCTCACCAG</td>
<td>60</td>
</tr>
<tr>
<td>f33</td>
<td>AAGTGGGCAAGACTGTTGAG</td>
<td>60</td>
<td>r33</td>
<td>CGTAGCTCGAGATGGACCT</td>
<td>59</td>
</tr>
<tr>
<td>f34</td>
<td>CACAGCTTATCTGGAAGTGCGTA</td>
<td>60</td>
<td>r34</td>
<td>GGCCTGTTAGTACCAAGCA</td>
<td>60</td>
</tr>
<tr>
<td>f35</td>
<td>GGTGCTTTTCATCGGTGAT</td>
<td>59</td>
<td>r35</td>
<td>GACTCGGGCATTATTCTTAG</td>
<td>59</td>
</tr>
<tr>
<td>f36</td>
<td>GTGATATTTTGGTGGCGACT</td>
<td>59</td>
<td>r36</td>
<td>GGATCATGTTCCCCGACCA</td>
<td>60</td>
</tr>
<tr>
<td>f37</td>
<td>TGTTGGGGCATGCTCATAAA</td>
<td>59</td>
<td>r37</td>
<td>CAGGAACACATTTGGCAGTG</td>
<td>60</td>
</tr>
<tr>
<td>Primer name</td>
<td>Primer sequence</td>
<td>TM</td>
<td>Primer name</td>
<td>Primer sequence</td>
<td>TM</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------</td>
<td>----</td>
<td>-------------</td>
<td>----------------</td>
<td>----</td>
</tr>
<tr>
<td>f5</td>
<td>CTCCCCCGAGGTTCTTTG</td>
<td>59</td>
<td>r5</td>
<td>CCGGGCTCTCAGAAGCA</td>
<td>59</td>
</tr>
<tr>
<td>f6</td>
<td>GGTGATGGTATAAGGACTTTTATCAAGA</td>
<td>59</td>
<td>r6</td>
<td>CATCAGCTTTTCTGCTAGGTTG</td>
<td>59</td>
</tr>
<tr>
<td>f7</td>
<td>TTAAATGTATCAAGGAGTGAAA</td>
<td>59</td>
<td>r7</td>
<td>CAGCAATCTTTCTGCTGTTG</td>
<td>58</td>
</tr>
<tr>
<td>f8</td>
<td>TGGGTTGCTAATCTGCTG</td>
<td>59</td>
<td>r8</td>
<td>AGGGGACAAATAGGACTTTG</td>
<td>61</td>
</tr>
<tr>
<td>f9</td>
<td>GCAACTGAAGGAAGGAAGAACC</td>
<td>58</td>
<td>r9</td>
<td>GAGAAAATCCCGTGTTGG</td>
<td>59</td>
</tr>
<tr>
<td>f10</td>
<td>CTCTAGAGGAGGCTCTCTCTTCTCT</td>
<td>59</td>
<td>r10</td>
<td>CACCAATCTTTAACCAACCTCTTG</td>
<td>59</td>
</tr>
<tr>
<td>f11</td>
<td>AAAAACCTATGACCCATTCTCA</td>
<td>58</td>
<td>r11</td>
<td>ATGTTCCAGGCTCAACCAAT</td>
<td>59</td>
</tr>
<tr>
<td>f12</td>
<td>CCGCTGCTACTGCTCTGTTG</td>
<td>59</td>
<td>r12</td>
<td>GAAAGCAAGCCCTAGTTAAA</td>
<td>57</td>
</tr>
<tr>
<td>f13</td>
<td>GAAAGGATCTGGCATGCTGCTG</td>
<td>60</td>
<td>r13</td>
<td>TCAAGCAACTGTTGAAGCTCTG</td>
<td>59</td>
</tr>
<tr>
<td>f14</td>
<td>TGGGAGGGCTAATCAAATTC</td>
<td>60</td>
<td>r14</td>
<td>GCAGTTACCAAGCTTTTGGCTG</td>
<td>60</td>
</tr>
<tr>
<td>f15</td>
<td>ATTCTTCCCAAGCCACACAA</td>
<td>60</td>
<td>r15</td>
<td>GGTGTATCCATGAAGACCTCGT</td>
<td>59</td>
</tr>
<tr>
<td>f16</td>
<td>AACAACCAACCCACCAAGG</td>
<td>59</td>
<td>r16</td>
<td>AGCCACCTGCAAGGAGGAT</td>
<td>59</td>
</tr>
<tr>
<td>f17</td>
<td>GGAAGGATCTGGCATGCTGCTG</td>
<td>60</td>
<td>r17</td>
<td>ACAGGGTCTTTCATTTTTCCA</td>
<td>58</td>
</tr>
<tr>
<td>f18</td>
<td>ATGTTCCAGGCTCAACCAAT</td>
<td>59</td>
<td>r18</td>
<td>TTGCGTTTGTCATTCTCAGG</td>
<td>59</td>
</tr>
<tr>
<td>f19</td>
<td>CCTCCAAAAGGAAGGACAA</td>
<td>59</td>
<td>r19</td>
<td>CCGGGACATACCATCCACTTC</td>
<td>60</td>
</tr>
<tr>
<td>f20</td>
<td>GCAACTGAAGGAAGGAAGAACC</td>
<td>58</td>
<td>r20</td>
<td>TTTCCCCAAACCCAGTGTAA</td>
<td>60</td>
</tr>
<tr>
<td>f21</td>
<td>GGAAGGGCTAATGAGGACTTTT</td>
<td>60</td>
<td>r21</td>
<td>GTCGTTTGTCATTCTCAGG</td>
<td>59</td>
</tr>
<tr>
<td>f22</td>
<td>TGGGAGGGCTAATCAAATTC</td>
<td>60</td>
<td>r22</td>
<td>GCAGTTACCAAGCTTTTGGCTG</td>
<td>60</td>
</tr>
<tr>
<td>f23</td>
<td>ATTCTTCCCAAGCCACACAA</td>
<td>60</td>
<td>r23</td>
<td>GGTGTATCCATGAAGACCTCGT</td>
<td>59</td>
</tr>
<tr>
<td>f24</td>
<td>AACAACCAACCCACCAAGG</td>
<td>59</td>
<td>r24</td>
<td>AGCCACCTGCAAGGAGGAT</td>
<td>59</td>
</tr>
<tr>
<td>f25</td>
<td>GGAAGGATCTGGCATGCTGCTG</td>
<td>60</td>
<td>r25</td>
<td>ACAGGGTCTTTCATTTTTCCA</td>
<td>58</td>
</tr>
<tr>
<td>f26</td>
<td>ATGTTCCAGGCTCAACCAAT</td>
<td>59</td>
<td>r26</td>
<td>TTGCGTTTGTCATTCTCAGG</td>
<td>59</td>
</tr>
<tr>
<td>f27</td>
<td>CCTCCAAAAGGAAGGACAA</td>
<td>59</td>
<td>r27</td>
<td>CCGGGACATACCATCCACTTC</td>
<td>60</td>
</tr>
<tr>
<td>f28</td>
<td>GCAACTGAAGGAAGGAAGAACC</td>
<td>58</td>
<td>r28</td>
<td>TTTCCCCAAACCCAGTGTAA</td>
<td>60</td>
</tr>
<tr>
<td>f29</td>
<td>GGAAGGATCTGGCATGCTGCTG</td>
<td>60</td>
<td>r29</td>
<td>GTCGTTTGTCATTCTCAGG</td>
<td>59</td>
</tr>
<tr>
<td>f30</td>
<td>TGGGAGGGCTAATCAAATTC</td>
<td>60</td>
<td>r30</td>
<td>GCAGTTACCAAGCTTTTGGCTG</td>
<td>60</td>
</tr>
<tr>
<td>f31</td>
<td>ATTCTTCCCAAGCCACACAA</td>
<td>60</td>
<td>r31</td>
<td>GGTGTATCCATGAAGACCTCGT</td>
<td>59</td>
</tr>
<tr>
<td>f32</td>
<td>AACAACCAACCCACCAAGG</td>
<td>59</td>
<td>r32</td>
<td>AGCCACCTGCAAGGAGGAT</td>
<td>59</td>
</tr>
<tr>
<td>f33</td>
<td>GGAAGGATCTGGCATGCTGCTG</td>
<td>60</td>
<td>r33</td>
<td>ACAGGGTCTTTCATTTTTCCA</td>
<td>58</td>
</tr>
<tr>
<td>f34</td>
<td>ATGTTCCAGGCTCAACCAAT</td>
<td>59</td>
<td>r34</td>
<td>TTGCGTTTGTCATTCTCAGG</td>
<td>59</td>
</tr>
<tr>
<td>f35</td>
<td>CCTCCAAAAGGAAGGACAA</td>
<td>59</td>
<td>r35</td>
<td>CCGGGACATACCATCCACTTC</td>
<td>60</td>
</tr>
<tr>
<td>f36</td>
<td>GCAACTGAAGGAAGGAAGAACC</td>
<td>58</td>
<td>r36</td>
<td>TTTCCCCAAACCCAGTGTAA</td>
<td>60</td>
</tr>
<tr>
<td>f37</td>
<td>GGAAGGATCTGGCATGCTGCTG</td>
<td>60</td>
<td>r37</td>
<td>GTCGTTTGTCATTCTCAGG</td>
<td>59</td>
</tr>
<tr>
<td>f38</td>
<td>TGGGAGGGCTAATCAAATTC</td>
<td>60</td>
<td>r38</td>
<td>GCAGTTACCAAGCTTTTGGCTG</td>
<td>60</td>
</tr>
</tbody>
</table>

*TM* = Melting temperature

Table 3: Primer sequences and TM’s used for both PCR and sequencing. Sequences are presented 5’ to 3’
2.11. Sequencing

Sequencing was carried out by the Allan Wilson centre for genomic sequencing (Allan Wilson Centre, Massey University, PO Box 11-222, Palmerston North, New Zealand http://www.allanwilsoncentre.ac.nz/AWCGSintro.htm). To prepare the PCR products for sequencing a Qiagen PCR purification kit was used to remove any substances that might inhibit the sequencing reaction. The protocol used QIAquick centrifuge columns. The DNA was first put into the column and then spun to bind it to the stationary phase. Next washing buffer was applied to the column and spun through to remove any impurities in the DNA. Finally the purified DNA was eluted in a clean collection tube.

The PCR product was then mixed with the relevant primer and milliQ H2O in the following concentrations: 2ng/100bp/15 µl of PCR product, 3.2 picomoles of primer, and then filled to 15 µl using milliQ H20.

The samples underwent full sequencing at the Allan Wilson centre. A full sequence run involves labelling of their products using cycle sequencing PCR, sequencing reaction cleanup using the CleanSeq magnetic bead system to remove unincorporated fluorescent ddNTPs, capillary analysis on the ABI3730 DNA Analyzer, and the analysis and uploading of sequencing data to the AWCGS server for customer retrieval.

2.12. Kidney histology

Only kidneys collected in 2009 were suitable for histological preparation, this was due to the fact that all older kidney tissue had been stored at -20 for several years without fixation.

Samples to be used for histology were fixed in 10% neutral buffered formalin (NBF) overnight. The tissue was sampled from several areas of the kidney where possible. Ideally samples were taken from the medulla, the cortex, and the papilla; however the amount of cystic tissue made identifying the different regions of the kidney difficult at times.

From each block multiple 4 micrometer sections were taken. Each block had 2 slides stained with H&E, 1 slide stained with Masson trichrome, and 8 sections on SF+ slides for use in Immunohistochemistry. Sectioning and staining was performed by Mandy Wilson of the University of Otago histology department.
2.13. Mutation analysis software

Although the sequence variants cannot be confirmed as causative at this time, an idea of how damaging they are can be obtained using bioinformatic tools to test how well the substitutions would be tolerated. These tools operate on protein sequence not nucleotide sequence, so the polymorphisms first needed to be translated. Several programmes were used to evaluate if the sequence variants identified were likely to be deleterious, as shown in Chan, 2008.

The Programmes used were A-GVGD, SIFT, and Polyphen. A-GVGD, from the International Centre for Research on Cancer (http://agvgd.iarc.fr/), involves the input of multiple aligned sequences of your gene of interest from several different species and then stating the substitutions of interest. The programme then assigns two scores, GV and GD which then places the substitution in a class, which in turn assigns a risk estimate to the mutation. GV is a measure of the observed biochemical variation in a given position and GD is a measure of the biochemical difference between the mutant and the normal variation established by GV. Thus if GV is a low number it means the variation for a given amino acid position is small, and if GD is equal to 0 then it falls within the normal variation making the mutation neutral. Alternately if the GD is a high number it means there is a great biochemical difference between the mutation and the normal variation, thus increasing the chances that the mutation is deleterious. Values of GV and GD were used to assign the substitution into a class with an attached risk estimate.

SIFT is a collection of tools from the J. Craig Venter institute (http://sift.jcvi.org/) for prediction of how amino acids substitutions affect protein function. SIFT has several input options to choose from allowing you to analyse multiple proteins, a single sequence by itself, or multiple related protein sequences. In order to get the best result out of the programme “SIFT related sequences” was used, this requires the input of multiple sequences and also the input of the substitutions of interest. The output for SIFT consists of a table showing each position in the sequence and then stating what amino acids are predicted to be tolerated and what ones would not be tolerated.

Polyphen is a tool which predicts the impact of polymorphisms on protein structure and function. Although it is designed for use on human proteins it is still able to be used for other animals. Polyphen is very thorough in its analysis of a SNP. The three main steps in its analysis are 1) sequenced based characterisation of the substitution site 2) Calculation
of a position specific score called the PSIC and 3) Calculation of structural parameters and
contacts.

In the first step the sequence that contains the SNP and the surrounding area are compared
to the UniProt database to see if they are located in sites with any of the following
annotations:

- DISULFID, THIOLEST, THIOETH bond or
- BINDING, ACT_SITE, LIPID, METAL, SITE, MOD_RES, CARBOHYD,
  SE_CYS site

Any hits are memorised by Polyphen for later comparison to any 3D models of the protein
that may be available.

Next the position specific independent counts (PSIC) score is calculated. This is done by
first identifying all sequences that share homology with the query sequence. The resulting
multiple alignment is used by the PSIC software to calculate the profile matrix. Elements
of the matrix (profile scores) are logarithmic ratios of the likelihood of a given amino acid
occurring at a particular position to the likelihood of this amino acid occurring at any
position (background frequency). Polyphen then computes the absolute value of the
difference between profile scores of both allelic variants in the polymorphic position. If the
difference between profile scores is large it indicates that the substitution is found rarely or
not at all across the sequences studied.

Finally the amino acid substitution is mapped to the known 3D structure of the protein.
This reveals if the replacement is likely to destroy the hydrophobic core of a protein,
electro’static interaction, interactions with ligands, or any other important feature of a
protein. If the 3D structure of a protein is not known it is possible to use the structure of
homologous proteins.

After these steps are complete Polyphen uses the results to make predictions about the
affect of the amino acid sequence variant. The predictions are as follows:

- **probably damaging**, i.e., it is with high confidence supposed to affect protein
  function or structure
- **possibly damaging**, i.e., it is supposed to affect protein function or structure
- **benign**, most likely lacking any phenotypic effect
- **unknown**, when in some rare cases, the lack of data do not allow Polyphen to
  make a prediction
Chapter 3: Results

3.1. Introduction

The purpose of this project was to identify the genetic cause of polycystic kidney disease (PKD) in a flock of New Zealand sheep. Additionally a more in-depth description of the associated phenotype was sought.

In order to aid the search for the causative mutation the SNP50 bead chip from Illumina was used to genotype genomic DNA samples from flock B and its descendants. This was useful as the list of genes associated with ARPKD is expansive and the gene that was eventually identified would have been overlooked in favour of many other genes and may not have been identified in a reasonable amount of time.

Given the complex and incomplete nature of the sheep pedigree it was not feasible for linkage mapping to be performed. A high degree of father daughter breeding takes place in flocks of sheep and many pedigree programmes are not set up to deal with this. However the main problem that we encountered with animals that were genotyped is the deficiency of information about the parents of many of the affected lambs, meaning an accurate pedigree could not be set up. Additionally, for the animals for which full information was available, it was often the case that samples from one or both of their parents was unavailable.

Given that linkage analysis was unable to be performed, it was decided that homozygosity mapping would be the best option. Using the data from the SNP50 beadchip, areas of concordant homozygosity were identified. These were considered areas of interest when searching for putative causative genes. As the disease is inherited in a recessive manner it was assumed that the causative gene would lie in a homozygous region that was shared by all affected animals.

Once identified, any areas of interest were examined using the sheep genome browser to find genes. Information about these genes was then searched to see if any link to the disease phenotype seen in the lambs could be established. PCR and sequencing of the gene found to have a link was carried out in an attempt to identify sequence variants.

Next an attempt to establish a correlation between any sequence variants and the inheritance of the disease phenotype was carried out. This involved the sequencing of
several affected animals, unrelated wild-types, and known carriers. Finally, the effect of
the sequence variants on the function of the protein was assessed using several different
types of polymorphism analysis software.

3.2. Animals

Breeding of flock B and it descendants was carried out as described in section 2.3.,
(pedigrees from these can be found in appendix C). Genomic DNA samples were obtained
from the resultant offspring of the breeding and from the rams and ewes upon death
following the method outlined in section 2.4.

The animals that were selected for genotyping are detailed in the table below.

<table>
<thead>
<tr>
<th>Name</th>
<th>Status</th>
<th>Father</th>
<th>Mother</th>
<th>Sex</th>
<th>Breed</th>
</tr>
</thead>
<tbody>
<tr>
<td>98-8587</td>
<td>Carrier</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Male</td>
<td>Coopworth</td>
</tr>
<tr>
<td>198/190</td>
<td>Affected</td>
<td>98-8587</td>
<td>8001</td>
<td>Male</td>
<td>Coopworth</td>
</tr>
<tr>
<td>07/498</td>
<td>Affected</td>
<td>06-5</td>
<td>01-43</td>
<td>Male</td>
<td>Coopworth</td>
</tr>
<tr>
<td>07/497</td>
<td>Affected</td>
<td>06-5</td>
<td>01-43</td>
<td>Male</td>
<td>Coopworth</td>
</tr>
<tr>
<td>95/97</td>
<td>Affected</td>
<td>98-8587</td>
<td>01-3</td>
<td>Male</td>
<td>Coopworth</td>
</tr>
<tr>
<td>06-10</td>
<td>Affected</td>
<td>98-8587</td>
<td>01-40</td>
<td>Male</td>
<td>Coopworth</td>
</tr>
<tr>
<td>07/195</td>
<td>Affected</td>
<td>06-26</td>
<td>01-23</td>
<td>Female</td>
<td>Coopworth</td>
</tr>
<tr>
<td>Ewe 4</td>
<td>Affected</td>
<td>Unknown</td>
<td>Ewe 4</td>
<td>Male</td>
<td>Coopworth</td>
</tr>
<tr>
<td>Ewe 5</td>
<td>Affected</td>
<td>Unknown</td>
<td>Ewe 5</td>
<td>Male</td>
<td>Coopworth</td>
</tr>
<tr>
<td>Ewe 7</td>
<td>Affected</td>
<td>Unknown</td>
<td>Ewe 7</td>
<td>Unknown</td>
<td>Coopworth</td>
</tr>
<tr>
<td>Ewe 8</td>
<td>Affected</td>
<td>Unknown</td>
<td>Ewe 8</td>
<td>Unknown</td>
<td>Coopworth</td>
</tr>
<tr>
<td>Ewe 9</td>
<td>Affected</td>
<td>Unknown</td>
<td>Ewe 9</td>
<td>Male</td>
<td>Coopworth</td>
</tr>
<tr>
<td>K3 lamb A</td>
<td>Affected</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Coopworth</td>
</tr>
<tr>
<td>K3 lamb B</td>
<td>Affected</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Coopworth</td>
</tr>
<tr>
<td>K4 lamb</td>
<td>Affected</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Coopworth</td>
</tr>
</tbody>
</table>

Table 3.1. Details of the animals that were sent for genotyping using the Ovine SNP50
beadchip.

The majority of the animals selected for genotyping were affected meaning they were born
with the PKD phenotype. One carrier was also sent. This carrier was able to act as a
negative control for homozygosity mapping. If this carrier was homozygous for a putative
region, then that region could be ruled out as causative.
3.3. DNA isolations

A large number of DNA samples were made fresh from historical frozen tissue using the Promega wizard genomic DNA kit. All DNA from the MKS flock that was used for SNP50 BeadChip hybridization, PCR, and sequencing was isolated using this method.

Table 3.2 details the DNA samples selected to be sent to Illumina for hybridization to the SNP50 ovine beadchip. For the purpose of this work an A260/280 ratio and gel electrophoreses of the DNA was used to assess quality. An A260/280 of 1.80 or higher was deemed acceptable as this indicates that the nucleic acid sample has little contamination from protein. Upon electrophoreses of the extracted DNA it should resolve as a single band with limited streaking.

<table>
<thead>
<tr>
<th>Name</th>
<th>DNA concentration (ng/μL)</th>
<th>A260/280</th>
<th>Gel band</th>
</tr>
</thead>
<tbody>
<tr>
<td>98-8587</td>
<td>53.6</td>
<td>1.95</td>
<td></td>
</tr>
<tr>
<td>198/190</td>
<td>79.71</td>
<td>1.80</td>
<td></td>
</tr>
<tr>
<td>07/498</td>
<td>72.21</td>
<td>1.98</td>
<td></td>
</tr>
<tr>
<td>07/497</td>
<td>77.4</td>
<td>1.91</td>
<td></td>
</tr>
<tr>
<td>95/97</td>
<td>93.41</td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>06-10</td>
<td>122.2</td>
<td>2.01</td>
<td></td>
</tr>
<tr>
<td>07/195</td>
<td>126.06</td>
<td>1.95</td>
<td></td>
</tr>
<tr>
<td>Ewe 4 Lamb</td>
<td>187.69</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td>Ewe 5 Lamb</td>
<td>62.1</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td>Ewe 7 Lamb</td>
<td>78.78</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td>Ewe 8 Lamb</td>
<td>97.81</td>
<td>1.89</td>
<td></td>
</tr>
<tr>
<td>Ewe 9 Lamb</td>
<td>94.98</td>
<td>1.87</td>
<td></td>
</tr>
<tr>
<td>K3 lamb A</td>
<td>129.22</td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>K3 lamb B</td>
<td>62.14</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>K4 lamb</td>
<td>91.55</td>
<td>2.07</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2. DNA concentration, 260/280 ratio, and gel electrophoresis bands of the DNA samples used for SNP50 beadchip genotyping. In combination these were used to assess the quality of the DNA extractions.
3.4. SNP50 BeadChip genotyping

Of the 15 samples that were sent to Illumina, 14 were able to be used on the SNP50 bead chip. The quality of the 15th sample, 06-10, was not good enough to be effectively hybridized to the SNP chip. The samples were hybridized in conjunction with 2866 samples from AgResearch, the success rates from this hybridization are summarised below in table 3.3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th># Successful</th>
<th># Possible</th>
<th>Success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample success rate</td>
<td>2,839</td>
<td>2,880</td>
<td>98.58%</td>
</tr>
<tr>
<td>Locus success rate</td>
<td>54,241</td>
<td>59,454</td>
<td>91.32%</td>
</tr>
<tr>
<td>Genotypes (Callrate)</td>
<td>151,658,204</td>
<td>153,030,617</td>
<td>99.10%</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>2,029,132</td>
<td>2,029,132</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 3.3. Success rates from the hybridization of the samples sent to Illumina for genotyping. The sample success rate is how many of the total samples sent were able to be hybridized. The locus success rate is how many of the total SNPs were able to be used. The genotypes is the number of SNP’s from all animals that worked. The reproducibility is the rate at which the results can be reproduced.

The output given to us from Illumina came in the form a large text file listing each SNP and the alternate bases that each sheep had in that position, designated allele A or allele B, as seen below in figure 3.1.

![Figure 3.1](image-url)
the SNP position (Base 1 and Base 2) the major and minor allele calls (Allele 1 and Allele 2), the chromosome the SNP is on (Chr) and the position on the chromosome (Position).

This figure shows the name of the SNP as well as the chromosome it is on and its position on that chromosome. The sample that is being looked at is listed and is identified by the number at the end. The columns labelled Base 1 and Base 2 show the base call at the SNP, and the columns labelled Allele 1 and Allele 2 show this in the form of major and minor alleles represented as A and B respectively.

Interpreting the raw data while it was in this format would have been very difficult so with the help of Dr. David Markie, (Department of Pathology, Dunedin School of Medicine) the data was manipulated in an excel spreadsheet to give a visual representation of the degree of homozygosity on each chromosome and highlight areas of concordant homozygosity. This was achieved by instructing excel to change the base calls into binary, where a homozygous call equals 1 and a heterozygous call equals 0. Homozygous base calls were highlighted in red as seen in figure 3.2. Next SNPs that had concordant homozygosity across all animals were assigned a 1 and highlighted in red, and finally regions where this concordant homozygosity was 100% of the SNP window size were marked as yellow, giving a region of interest.

Figure 3.2. In this figure the red vertical bars represent homozygous regions in each animal, and the yellow bar at the far right represents a run of concordant homozygosity. Each column is a different sheep and each row is a different SNP. In the enlarged box on the left we see a close up of several rows, in each row we see from left to right, the name of the SNP, the chromosome it is located on, and its physical location in base pairs. In the second enlarged box it is possible to see the alleles of a selection of the animals. The particular area that has been enlarged is showing a region of concordant homozygosity.
3.5. Homozygosity mapping

From the values derived for concordant homozygosity, graphs of each chromosome could be plotted. These illustrated the level of autozygosity across the genome, and allowed for regions of concordant homozygosity to be easily identified. Across the majority of the genome the level of concordant homozygosity was usually between 20% and 60% with occasionally peaks reaching higher. Only peaks that reached 100% concordant homozygosity were of interest in identifying putative causative regions.

Figure 3.3. Excel graphs representing the degree of concordant homozygosity on chromosomes 4, 9, and 11
The graphs in figure 3.3 represent chromosomes 4, 9, and 11 respectively. On the x axis of the graphs is the physical distance along the chromosome in mega bases and on the y axis is the concordant homozygosity score. The window for the number of SNPs that all animals must be concordantly homozygous for is set at 10, as indicated by the (10) next to the concordant homozygosity key. The concordant homozygosity score shows the percentage of the animals that are concordantly homozygous for the same 10 SNPs in any given region of the chromosome. Increasing the SNP window size causes the regions of homozygosity on the graphs to be lost. The region on chromosome 4 is lost at 15 SNPs, the region on chromosome 9 is lost at 30 SNPs, and the region on chromosome 11 is lost at a window size of 20 SNPs.

As can be seen in the above graphs, only three areas of concordant homozygosity were identified in the entire genome giving an idea where to look in the genome for a possible causative gene. By looking at the SNPs that define the ends of each region of concordant homozygosity it is possible to find out the physical region, in base pairs, that the homozygosity spans. This was then put into the sheep genome browser (refer to section 2.9. for reference) allowing for a visualisation of the putative region and all the genes that were mapped to it.

### 3.6. Gene Identification

From the homozygosity mapping the areas identified as being of interest were as follows:

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Base Pairs</th>
<th>Identified or known Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>bp 72380844 to bp 72769785</td>
<td>3-hydroxyisobutyrate dehydrogenase.</td>
</tr>
<tr>
<td></td>
<td>(388941 bp long)</td>
<td>Even skipped homeobox1 (EVX1).</td>
</tr>
<tr>
<td>9</td>
<td>bp 87942007 to bp 89018996</td>
<td>TMEM67.</td>
</tr>
<tr>
<td></td>
<td>(1076989 bp long)</td>
<td>RNA binding motif protein 12B.</td>
</tr>
<tr>
<td>11</td>
<td>bp 27807356 to bp 29412465</td>
<td>60 genes were found in this area (Full list in appendix D)</td>
</tr>
<tr>
<td></td>
<td>(1605109 bp long)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4. Locations of identified regions of homozygosity and the genes associated with them
The two genes that were found on the chromosome 4 regions were 3-hydroxyisobutyrate dehydrogenase and Even skipped homeobox1 (EVX1). Neither of these genes have any known association to polycystic kidney disease.

Two genes were also found in the area of interest on chromosome 9. Information about RNA binding motif protein 12B is limited and no association with PKD was found. However TMEM67 is recognised as playing a role in several syndromes that have ARPKD as a characteristic.

Of the 60 genes found in the region of interest on chromosome 11 the only gene of interest was TMEM95 which is known to have an association to adult onset PKD and liver injury. However the adult onset nature of the PKD made TMEM95 an unlikely candidate due to the early onset of disease presentation in affected lambs.

Of all the genes in these areas only one was a strong candidate gene. This was TMEM67 found in the region on chromosome 9 as seen in figure 3.4.

Figure 3.4. Virtual sheep genome image showing the region of concordant homozygosity of chromosome 9 extending from base pair 87942007 to base pair 89018996 aligned with the human genome reference sequence and the Ovine SNP50 BeadChip SNPs.
When looking at the relevant information about TMEM67 it was found that it is responsible for several recessively inherited syndromes such as Meckel-Gruber and Joubert syndrome, both diseases which have ARPKD as a characteristic.

### 3.7. PCR

Using sequence obtained from AgResearch, primer sets were designed to amplify each exon. The majority of the exons in TMEM67 were able to be amplified by PCR. Of 28 exons 26 were amplified. In addition a considerable amount of intronic sequence was also obtained. The results of the PCR of the TMEM67 exons is summarised in table 3.5. The table details the exon being sequenced, the primers used to amplify it, the predicted size of the PCR product, and the size shown on the gel band. Pictures of the gel bands can be found in appendix E.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Primers</th>
<th>Predicted size</th>
<th>Gel band Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TMEM67-fR, TEME67-rR</td>
<td>918</td>
<td>900</td>
</tr>
<tr>
<td>2</td>
<td>TMEM67-intron1F, TMEM67-intron2R</td>
<td>451</td>
<td>450</td>
</tr>
<tr>
<td>3</td>
<td>TMEM67-fQ, TMEM67-rQ</td>
<td>311</td>
<td>300</td>
</tr>
<tr>
<td>4</td>
<td>TMEM67-fL, TMEM67-rL</td>
<td>570</td>
<td>550</td>
</tr>
<tr>
<td>5</td>
<td>TMEM67-fK, TMEM67-rK</td>
<td>745</td>
<td>650</td>
</tr>
<tr>
<td>6</td>
<td>TMEM67-fJ, TMEM67-rJ</td>
<td>407</td>
<td>400</td>
</tr>
<tr>
<td>7</td>
<td>TMEM67-f43, TMEM67-r43</td>
<td>900</td>
<td>850</td>
</tr>
<tr>
<td>8</td>
<td>TMEM67-f41, TMEM67-r41</td>
<td>966</td>
<td>1000</td>
</tr>
<tr>
<td>11</td>
<td>TMEM67-f39, TMEM67-r39</td>
<td>968</td>
<td>1000</td>
</tr>
<tr>
<td>12</td>
<td>TMEM67-f34, TMEM67-r34</td>
<td>905</td>
<td>900</td>
</tr>
<tr>
<td>13</td>
<td>TMEM67-fH, TMEM67-rH</td>
<td>567</td>
<td>550</td>
</tr>
<tr>
<td>14</td>
<td>TMEM67-f33, TMEM67-r33</td>
<td>754</td>
<td>700</td>
</tr>
<tr>
<td>15</td>
<td>TMEM67-fG, TMEM67-rG</td>
<td>549</td>
<td>500</td>
</tr>
<tr>
<td>16</td>
<td>TMEM67-f28, TMEM67-r28</td>
<td>351</td>
<td>400</td>
</tr>
</tbody>
</table>
Table 3.5. Summary of the PCR amplification of the exons of the ovine TMEM67 gene.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Exon Name</th>
<th>Forward Base</th>
<th>Reverse Base</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>TMEM67-f26, TMEM67-r26</td>
<td>757</td>
<td>700</td>
</tr>
<tr>
<td>18</td>
<td>TMEM67-fM, TMEM67-rM</td>
<td>610</td>
<td>650</td>
</tr>
<tr>
<td>19</td>
<td>TMEM67-fF, TMEM67-rF</td>
<td>506</td>
<td>500</td>
</tr>
<tr>
<td>20</td>
<td>TMEM67-fE, TMEM67-rE</td>
<td>650</td>
<td>650</td>
</tr>
<tr>
<td>21</td>
<td>TMEM67-fD, TMEM67-rD</td>
<td>395</td>
<td>400</td>
</tr>
<tr>
<td>22</td>
<td>TMEM67-fO, TMEM67-rO</td>
<td>360</td>
<td>350</td>
</tr>
<tr>
<td>23</td>
<td>TMEM67-f13, TMEM67-r13</td>
<td>902</td>
<td>850</td>
</tr>
<tr>
<td>24</td>
<td>TMEM67-fC, TMEM67-rC</td>
<td>679</td>
<td>650</td>
</tr>
<tr>
<td>25</td>
<td>TMEM67-fC, TMEM67-rC</td>
<td>679</td>
<td>650</td>
</tr>
<tr>
<td>26</td>
<td>TMEM67-fB, TMEM67-rB</td>
<td>858</td>
<td>850</td>
</tr>
<tr>
<td>27</td>
<td>TMEM67-fA, TMEM67-rA</td>
<td>561</td>
<td>550</td>
</tr>
<tr>
<td>28</td>
<td>TMEM67-fP, TMEM67-rP</td>
<td>383</td>
<td>400</td>
</tr>
</tbody>
</table>

3.8. Sequencing

The 26 exons that were able to be amplified using PCR were sequenced using the same primers as were used to carry out the PCR reaction. Quality sequence was able to be obtained in most cases as seen in the chromatograms included in appendix D. Exons 9 and 10 were unable to be amplified despite several attempts and using several different primers. In most cases where a heterozygous base was called or the peaks were indistinct the base was able to be resolved by looking at the sequence generated from using the reverse primer.

Looking through the chromatograms it is seen that there are two heterozygous base calls in exon 20 as seen in figure 3.5. These were present in both forward and reverse primers. Chromatograms, as well as DNA and amino acid sequence for all exons can be found in appendix F.
Exon 20 (primer set “E”) putative mutation

When these sequence variants were identified sequencing was carried out on other animals including affected lambs, wild-type lambs, and carriers to confirm that one, or both, of the polymorphisms were inherited in the correct manner. The results are seen in figure 3.6.

Figure 3.5. Chromatograms of a portion of exon 20 in a carrier, affected, and wild type sheep showing two substitution mutations.
Figure 3.6. Alignment of several carrier (A, B, C), wild type (I, J), and affected animals in (F, G, E, D) the region of interest.

From this figure we see that the inheritance of both sequence variants does seem to follow a pattern consistent with the inheritance of the disease. The first polymorphism is found at base pair 79 in exon 20; it is represented as a W in the carriers. In wild type animals the base is called as a thymine while in the affected animals it is called as an adenine. The second polymorphism is found at base pair 97 and is identified by a K. Wild types again have a thymine in this position with the mutants possessing a guanine.

In addition to the two wild types sequenced above, 30 more unrelated wild type sheep were also sequenced in this area. None of these wild type sheep had either mutation.

3.9. Polymorphism analysis

For the polyphen analysis only the wild type sheep data was entered, as the programme automatically aligns the query sequence against the nrdb database. The outputs for both polymorphisms are displayed below:
Figure 3.8. Polyphen outputs for both mutations.

As can be seen in the images above both polymorphisms were predicted to be probably damaging, which means the programme has high confidence that the variant will affect protein function. This prediction is based on alignments to other sequences found in the nrdb database. Both were also scored with very high PSIC scores indicating that these substitutions are rarely or never observed in this protein family.

For both A-GVGD and SIFT, sequences from sheep, rat (ID: NP_00101386.2), mouse (ID: NP_808529), cow (ID: XP_583257), human (ID: AAH32835), zebra fish (ID: XM_695882.4), and chicken (ID: XM_418334.2) were input in an attempt to stop conservation shared by closely related species from biasing or giving a false result. These sequences were all selected by the Polyphen program and so where used for A-GVGD and SIFT in order to limit variables between the programs.

When the sequences with the polymorphism were put into AGVGD the results were as follows:

<table>
<thead>
<tr>
<th>Substitution</th>
<th>GV</th>
<th>GD</th>
<th>Prediction</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>I640N</td>
<td>0.00</td>
<td>148.91</td>
<td>Class C65</td>
<td>Predicted Deleterious</td>
</tr>
<tr>
<td>I646S</td>
<td>0.00</td>
<td>141.80</td>
<td>Class C65</td>
<td>Predicted Deleterious</td>
</tr>
</tbody>
</table>

Table 3.6. Results from AGVGD detailing the name, the GV and GD scores, and the subsequent prediction. Substitution I640N refers to the first polymorphism and I646S refers to the second.
When we reference the GV, GD, and class scores back to the AGVGD risk assessment graph below we see that both polymorphisms lie in the high risk area meaning that AGVGD considers them both to be deleterious mutations.

When the same information was put into SIFT a similar result was obtained. At the positions of both polymorphisms it is seen that the amino acid substitution found in the mutant sheep is classed as not being tolerated meaning that SIFT considers both to be deleterious to the protein.

<table>
<thead>
<tr>
<th>Predict Not Tolerated</th>
<th>Position Seq Rep</th>
<th>Predict Tolerated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>636E 1.00</td>
<td>E</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>637W 1.00</td>
<td>W</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>638N 1.00</td>
<td>N</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>639E 1.00</td>
<td>E</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>640I 1.00</td>
<td>I</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>641Q 1.00</td>
<td>Q</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>642T 1.00</td>
<td>T</td>
</tr>
<tr>
<td>hw'dqng'rpecksyfmta</td>
<td>643V 1.00</td>
<td>I     V</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>644R 1.00</td>
<td>R</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>645K 1.00</td>
<td>K</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>646I 1.00</td>
<td>I</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>647N 1.00</td>
<td>N</td>
</tr>
<tr>
<td>wfmymichlqrvkednt</td>
<td>648P 1.00</td>
<td>g     a S P</td>
</tr>
<tr>
<td>whdgcnyeqrpks</td>
<td>649L 1.00</td>
<td>f     ma v i TL</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>650F 1.00</td>
<td>F</td>
</tr>
<tr>
<td>ywvtsrqpmlkhihgfedca</td>
<td>651Q 1.00</td>
<td>Q</td>
</tr>
</tbody>
</table>

Substitution at pos 640 from I to N is predicted to AFFECT PROTEIN FUNCTION with a score of 0.00.

Substitution at pos 646 from I to S is predicted to AFFECT PROTEIN FUNCTION with a score of 0.00.

Figure 3.7. An output from the SIFT programme showing the alignment and predictions for both mutations. The column labelled “PositionSeq” shows the amino acid position and the amino acid that is present in the reference sequence. The left hand side of the image shows the amino acids that are predicted to not be tolerated there, and the right hand side shows those that are.
As seen in figure 3.7 both the polymorphisms called I640N and I646S were predicted to affect protein function.

<table>
<thead>
<tr>
<th>SNP</th>
<th>A-GVGD</th>
<th>SIFT</th>
<th>Polyphen</th>
</tr>
</thead>
<tbody>
<tr>
<td>I638S</td>
<td>Predicted deleterious</td>
<td>Predicted to affect protein function</td>
<td>Predicted to be probably damaging</td>
</tr>
<tr>
<td>I644N</td>
<td>Predicted deleterious</td>
<td>Predicted to affect protein function</td>
<td>Predicted to be probably damaging</td>
</tr>
</tbody>
</table>

Table 3.6. Summary of the polymorphism analysis at both site by the three programmes used.

When sequences of this protein from a range of species are aligned a high level of conservation is found in both sequence variant positions. This conservation is even found in *Gallus Gallus* (Chicken), *Danio Rerio* (Zebra fish) and *Ornithohynexus anatinus* (Platypus) indicating that this stretch of sequence may play an important role in the normal functioning of the Meckelin protein.

Sheep           TYFVANEWNE IQTVRKINPL FQVLTLLFLLE E
Cow             TYFVANEWNE IQTVRKINPL FQVLTLLFLLE E
Rat             TYFVANEWNE IQTVRKINPL FQVLTLLFLLE E
Mouse           TYFVANEWNE IQTVRKINPL FQVLTLLFLLE E
Human           TYFVANEWNE IQTVRKINSL FQVLTLLFLLE E
Chicken         TYFIANEWNE IQTVRKINPL FQVLAFLLE E
Zebra fish      TYFVANEWNE IQTIRKINPT FQVMAFLLE E
Puffer fish     TYFVANEWNK IQTIRQSPFT FQIMAVLFE E
Possum          TYFVANEWNE IQTVRKINPL FQVLTLLFLLE E
Orang-utan      TYFIANEWNE IGTVRKINSL FQVLTLLFLLE E
Rhesus Macaque  TYFANVEWNE IGTVRKINSL FQVLTLLFLLE E
Marmoset        TYFVANEWNE IGTVRKINSL FQVFLFLE E
Platypus        STTAPNEWNE IGTVRKINPL FQVAVLLE E
Wild Boar       TYFVANEWNE IQTVRKINPL FQVLTLLFLLE E
Dog             TYFVANEWNE IQTVRKINPL FQILTVLLFLLE E

Figure 3.9. The segment of the Meckelin protein in which the two polymorphisms are found. I640N highlighted in yellow and I646S highlighted in red. From this we see that these amino acids are well conserved across many species.
The structure of Meckelin is yet to be resolved meaning it is not possible to know what role the identified sequence variants play. However, looking at the model proposed by Dawe et al., 2009 in figure 3.10 the sequence would appear to correspond to an intracellular loop.

Figure 3.10. Domain structure of meckelin as predicted by Dawes et al., 2009 showing signal peptide, cysteine-rich repeat region, three transmembrane domains, and a coiled-coil domain.

3.10 Tissue collection

Of the lambs born in 2009 from the PKD flock, 8 dissections were performed successfully. Five of these were identified as possible affected lambs in utero by the ultrasound scans and three were unaffected. These lambs were labelled 09/A, 09/B, 09/C, 09/D, 09/E, 09/F, 09/G, and 09/H respective to the order in which they were born. In addition to these, two unrelated wild types that were born in the same lambing season were also dissected to act as controls and labelled 09/I and 09/J.

The affected lambs, 09/D, 09/E, 09/F, 09/G, and 09/H, all had similar disease characteristics. When born they were often found to have distended abdomens or scrotums. Many died during the birthing process. Those that survived were euthanized quickly so as not to prolong any suffering they may have been experiencing.

Upon opening of the abdominal cavity of the affected lambs the cause of the distended abdomens was found to be massively enlarged kidneys. When the kidneys were sectioned they were found to be filled with a large number of micro cysts that often made up the majority of the kidneys mass. These cysts, as well as enlarging the kidneys, also made it
difficult to identify the different regions of the kidneys internal structure and likely possibly hinder the function.

Hepatic abnormalities were present in many of the animals, with both cystic and cirrhotic liver tissue being found. The extent of liver damage was variable between animals with some animals having little to no fibrosis and some having such extensive damage that the majority of the liver is fibrotic as seen in figure 3.13. In one case what appeared to be a secondary liver was found and as patterning abnormalities are often associated with ciliopathies it is plausible that this could have been caused by the mutation. However no other patterning problems were identified in any other animal, so this cannot necessarily be attributed to the disease.

The pancreas, bile ducts, and epididymis were usually highly cystic. The cysts were often larger and fewer than those found in the kidneys but still made up a significant portion of the organ.

All these findings were in agreement with the phenotype seen in affected lambs from flock A as described in Johnstone et al., 2005.

Figure 3.11. A new born lamb with an extremely distended abdomen. Upon opening the abdominal cavity it is seen that the kidneys are enlarged and occupy the majority of the available space.
Figure 3.12. (Left) A transverse section of an unaffected kidney (Right) A section of an affected kidney showing that cysts are pervasive throughout the kidney making it difficult to differentiate between the Cortex and Medulla.

Figure 3.13. As seen above, the livers of affected lambs often have a large amount of fibrous and scar tissue as well as some cystic tissue. It is also possible to see a part of a cystic bile duct just below the liver in the left picture (Arrow). The enlarged kidney of a new born lamb (top) compared to the kidney of a full grown Ram (bottom).

From these dissections a good representation of the phenotype from the MKS flock was obtained and found to be in agreement with what was described for flock B.
3.11. Histology

Haematoxylin and eosin staining was performed on sections taken from the kidneys of the lambs dissected in 2009. Several images from this are included below.

Figure 3.14. H and E staining of both wild type and mutant sheep kidneys. A) Wild type kidney section at 50 times magnification. B) Mutant kidney section at 50 times magnification. C) Wild type kidney at 100 times magnification. D) Mutant kidney at 100 times magnification.

The images above are typical of all the samples that underwent histology. The wild type slides give a good approximation of what a healthy kidney should look like. A large number of glomeruli are easily visible and the surrounding tissue is dense, unified, and organized. In contrast the mutant tissue is characterised by the large vacuous spaces caused by the cysts. The structure of the affected kidneys is greatly disturbed by the cysts, and the surrounding cells are compressed. On the interior of the cysts of image D it is possible to see a large amount of cell debris. The arrow on image B shows the presence of a glomeruli tuft in the cystic space, this demonstrates that this disease is able to induce cyst formation in the glomeruli.
Chapter 4: Discussion

The work presented in this thesis was performed with the intention of developing the PKD sheep as a model organism for autosomal recessive polycystic kidney disease. In order to achieve this, the phenotype of the animals had to be well established and the genotype had to be identified. In addition to these things it was important to assess the suitability of the PKD flock as a model for human disease. To achieve this, the suitability of sheep as a human disease model had to be assessed by looking at what ovine models are currently used. Additionally a comparison between the phenotype of the PKD sheep, the human disease, and the rodent disease models was required to see if the phenotype was a good representation of that seen in humans.

4.1. Identification of the genotype

4.1.1. SNP50 BeadChip

A key component of this project was the use of the SNP50 ovine BeadChip. SNPs by themselves are not overly informative with regard to identifying linkage markers of interest. This is due to there being only two alleles for each SNP. This is in contrast to markers such as restriction fragment length polymorphisms, which are able to have many different alleles allowing for linkage between a particular marker and the genetic feature of interest to be established. However when SNPs are put in a high enough density they are able to give a great deal of information, as a high density segment of SNPs becomes similar to a haplotype. The SNP50 ovine BeadChip put this to use by encompassing 50,000 SNPs across the sheep genome. Importantly, in choosing the SNPs to include on the chip AgResearch tried to space the SNPs evenly. The end result being that the SNP50 BeadChip is able to be highly informative in regards to linkage of genetic markers and phenotypes of interest.
4.1.2. Homozygosity mapping

As explained in the methods, linkage was unable to be performed due to lacking information about parentage, or lacking samples to complete a fully informative pedigree. Without a pedigree it is not possible to make predictions about recombination, which is an important step for linkage as it is part of what is required to calculate the logarithm of odds score. Although linkage would have been the preferred method for detecting regions of interest, homogygosity mapping is able to do the same thing. However the accuracy of homogygosity mapping may not be as high as the accuracy for linkage analysis. There is also no statistics associated with homogygosity mapping so it is impossible to know if the regions of interest you are looking at are real or are just artefacts of inbreeding. Due to this it is important to thoroughly scrutinize the list of candidate genes and be careful in which genes you choose to pursue.

When looking for regions of homogygosity it was possible to change the number of homogygous SNPs that would be called as an area of interest. The number chosen was important as changing it changed the number and size of the areas identified. If the window was set too low then many areas of homogygosity would have been found. However many of these would most likely be false positives that would just be small shared regions that very little recombination occurs in and are shared in all animals due to the level of inbreeding present in all flocks of sheep. This was especially likely in this flock as it has been thoroughly inbred for the purpose of consistently producing affected lambs. Alternatively if the window is set too high then no regions of homogygosity will be identified. Due to recombination events getting more common as you look at larger regions large stretches of concordant homogygosity are rare. For these reasons several different window sizes were tried, most commonly the window size was between 10-25 SNPs. At the bottom end of this size range all three areas identified in the results were called as areas of interest, as the window size was increased the region on chromosome 11 was lost. The regions on chromosomes 4 and 9 were lost at about the same time. Even though the region on chromosome 11 was lost with slightly higher window sizes it was still considered a region of interest as it was still a relatively large area of concordant homogygosity. In the results section the window size in the graphs is set to 10 SNPs this was done simply to give clear pictures of the homogygosity. Given the high level of inbreeding that has occurred in this particular flock of sheep it is not surprising that three regions of concordant homogygosity were found.
When performing the homozygosity analysis the animal 95/97 had to be left out as when it was included no regions of homozygosity were identified by the programme. But when we look at the SNP most closely related with the gene TMEM67 we find that 95/97 is, in fact, concordantly homozygous for the area in which this SNP lies. There are several reasons that 95/97 may have removed all areas of interest.

4.1.3. Sequencing

Given that TMEM67 was the best candidate gene found, sequencing was focused on it. Initially, attempts were made to try and amplify as much of the gene as possible for sequencing. However the 2009 paper by Otto et al indicated that mutations in this gene that cause the phenotypes we were seeing are usually either non-sense or missense. That is that they are usually located in coding regions and are mostly single nucleotide polymorphisms leading to either an amino acid substitute or a stop codon. This caused the sequencing to become focused on the exons of the TMEM67 gene. The initial sequence for this gene was obtained from AgResearch who are currently sequencing the sheep genome. However the sequence obtained was incomplete with large portions remaining unknown. Another potential problem was that the version of the sheep genome available to us was put together based upon the bovine genome. Although the genomes of the two species should be highly similar due to their evolutionary proximity, there are still going to be differences. When looking to map the exons to the sequence obtained from AgResearch it was noted that Ensemble only has 20 exons listed for the bovine sequence instead of the 28 listed for humans and most other animals. Because of this the human exons sequence was selected to act as the consensus sequence. To the original sequence obtained from AgResearch the exons 1, 2, 3, 7, 8, 12, 13, and 18 were unable to be mapped. In order for exons 1, 2, and 3 to be found additional upstream sequence had to be downloaded on top of the sequence given to us by AgResearch. However even with this additional sequence exon 2 was still absent. To find these missing exons then required sequencing of unknown fragments, primers were designed around the places most likely to contain the exons and sequencing was carried out. This additional sequence allowed for all the exons to be mapped to the gene. As was stated in the results section PCR products covering all the exons apart from exons 9 and 10 were successfully amplified and sequenced. Several different primers were made to try and amplify exons 9 and 10 but all are unsuccessful at this time. What makes amplifying exons 9 and 10 difficult is that they are mapped to a region with large amounts of unknown sequence; this means that the possible primers are
limited. Other than exons 9 and 10, the sequence from all other exons was of good quality. When the carrier animals were sequenced the only consistent heterozygous calls found were those seen in exon 20. Any other heterozygous call was resolved when additional sequencing was performed.

4.1.4. Polymorphisms

The polymorphisms are found at positions 79 and 97 in exon 20; this translates to positions 638 and 644 in the protein sequence. In the wild type sequence, the codon of the first mutation is ATT which codes for Isoleucine. This changes to AAT which codes for Asparagine. The second locus is also an ATT or Isoleucine in the wild types, this becomes AGT which codes for Serine. Because of this the mutations will be referred to as I638N and I644S, where I stands for isoleucine, N stands for Asparagine, and S stands for serine based on the IUPAC single letter abbreviations. Both I638N and I644S are seen to distribute in a pattern consistent with the disease phenotype. However, just distributing in the correct fashion does not mean that the polymorphisms are causative as they could just be in linkage with the actual causative mutation. For that matter it is possible that only one of the polymorphisms is causative and the other is in linkage, this could be probable given that the two mutations are located only 17 base pairs apart giving very little space for recombination to occur. Even if both mutations are just linked and not causative they are still able to act as a useful marker for genotyping the animals. Given the close proximity that the two sequence variants have to each other it is likely that they are both the result of the same mutational event.

To try and falsify the hypothesis that one or both of these mutations is causative, bioinformatic analysis was used to assess the effect that they would have on protein function. The 2008 paper “Interpreting missense mutations in Human TRIM5alpha by computational methods” by Philip A Chan outlined a method in which four computational tests were used to assess the effect that several polymorphisms were likely to have. The programs used were SIFT, Polyphen, A-GVGD, and average BLOSUM62 pairwise score. The conclusion of their work was that “Comparative sequence analysis offers a functional tool to analyze unknown nsSNPs”. In a previous paper Chan et al. showed that when all four of these programmes were in agreement they had a predictive value of 88.1%, meaning that although these methods do not use protein structural change they are still
very accurate at predicting the effect of SNPs. Because the methods used rely upon the evolutionary conservation of sequence to make predictions they are susceptible to bias if the variation in the input sequences isn’t great enough. To overcome this it is necessary to use sequences from organisms that are evolutionarily distant, it is for this reason that as many sequences as were available were used in the analysis. Unfortunately the Polyphen program, which selects its own sequences, limited the number of sequences for the TMEM67 gene to 9, and these were all from higher organisms including the zebra-fish, the short-tailed grey possum, sheep, and humans. This encompasses about 500 million years of evolutionary divergence between the fish and the higher mammals, and although the number of animals included was limited this cannot be over looked especially given that the sequence was conserved perfectly in both loci of interest across all animals.

4.2. TMEM67

The gene TMEM67 is 47kb long and encodes the protein Meckelin and has been identified as playing a role in many diseases. Meckelin has been predicted to be either a seven or a three transmembrane receptor protein with a large extracellular N-terminus containing a signal peptide and cystine rich domain. The intracellular C-terminus contains a coiled-coil domain. Crystallography to determine the structure of this protein is difficult due to it being associated with the cell membrane. Meckelin has been hypothesised to have several functions. It is known to play a role in centriole migration, which is an important step in ciliogenesis. To achieve this role, it is thought that Meckelin interacts with the Meckel-Gruber syndrome type 1 protein, MKS1, which localises to the centriole and the cell membrane. MKS1 would then provide a link with the protein serine/threonine kinase p160ROCK which is known to play a role in the formation and rearrangement of actin stress fibres. This interaction of Meckelin, MKS1, and p160ROCK then allows for the actin or tubulin rearrangements required for centriole migration.

Williams et al., (2010) found that knocking out TMEM67 in *C.elegans* caused elongated cilia, a phenotype that has been reported for MKS3 mutations in human, mice, and rats. In addition to the elongation of the cilia in mutant *C. elegans*, Williams et al., (2010) also noted the presence of multiple centrosomes and multiple cilia. They also found that TMEM67 and MKS1 interact with Nphp-1 and Nphp-4 to influence proper positioning, orientation, and formation of cilia. The authors putatively identified TMEM67 as acting
upstream of Nphp-4, as when both were knocked out the effect was not greater than when TMEM67 was knocked out by itself. Williams et al 2010 also came to the conclusion that the two pathways share similar functions allowing for some redundancy, and that the MKS1/MKS3 pathway has a partial role in the biology of the cilium membrane as opposed to the axoneme.

An addition role for TMEM67 was found by Wang et al 2009. This paper was looking at the mechanisms of the endoplasmic reticulum-associated degradation (ERAD) pathway which is involved in removal and degradation of non-folding proteins. When they stressed the ERAD pathway, either genetically by expressing non-folding proteins or pharmacologically, they found that expression of TMEM67 was increased 5 fold. Subsequent knockouts of TMEM67 in lung cells resulted in the build up of non-folding proteins that would normally be removed by the ERAD pathway, this build up eventually lead to formation of cytotoxic aggregates and activation of apoptosis. They concluded that the evidence supports a model in which TMEM67 links the ER luminal quality control machinery with the cytosolic degradation apparatus. Summarily, TMEM67 plays a role in both ciliogenesis and in the endoplasmic reticulum-associated degradation pathway. It has a putative function in protein trafficking in ERAD; this is in addition to its possible role of tethering transition fibres and remodelling of actin stress fibres in coordination with MKS1 and p160ROCK to allow for basal body migration.

4.3. TMEM67 associated diseases

As discussed in chapter 1 TMEM67/MKS3 has been found to be associated with many diseases, syndromes, and symptoms. Given its role in cilia formation and regulation this is not surprising as cilia are found in almost every cell type in the body, and when disrupted it becomes apparent just how crucial of an organelle they are to normal function.

The four main syndromes that MKS3 is associated with are Bardet-Biedl syndrome (BBS), Meckel-Gruber syndrome (MKS), Joubert syndrome (JS), and Nephronophthisis (NPHP).

BBS has a high level of phenotypic heterogeneity; the main symptoms associated with BBS are obesity, retinitis pigmentosa, polydactyly, mental retardation, hypogonadism, and renal dysfunction. The genetics of BBS is complex and thus far 15 genes have been associated with the disease. In a majority of cases the causative gene is BBS1 but there are many documented cases of BBS in which the disease only precipitates when an individual
inherits multiple defective alleles. The alleles that are inherited play a role in the presentation of the disease phenotype. MKS3 has been identified as a modifier in BBS. That means that when it is defective in association with one of the BBS genes it changes the phenotype, usually to something more damaging.

Meckel-Gruber syndrome is a ciliopathy that has had 6 loci associated with it; the two most well known genes are MKS1 and MKS3. MKS is most commonly identified by a triad of symptoms these being polydactyly, bilateral renal cyst dysplasia and liver defects. However there are additional symptoms that occur more rarely such as brain defects and gonadal malformations.

Another disease caused by defects in MKS3 is Joubert syndrome. The symptoms of JS closely resemble those of MKS; however it is associated with central nervous system defects to a greater extent than MKS.

Finally there is NPHP which is a very highly genetically heterogeneous disorder that has been shown to have an association with MKS3. It has been suggested that truncating mutations in MKS3 lead to either MKS or JS while missense mutations lead to the milder phenotype of NPHP. NPHP is characterised by cystic kidneys and hepatic dysfunction, and in rare cases situs inversus.

From looking at these syndromes we seen that defects in MKS3 are the cause of a wide range of symptoms including: CNS malformation, cystic kidneys, hepatic dysfunction, retinal degeneration, pancreas and biliary duct malformation, situs inversus, gonadal malformation, polydactyly, obesity, heart dysfunction, and diabetes. These symptoms are summarised along with the syndromes they are associated with in figure 4.2.

The Otto et al paper “hypomorphic mutations in meckelin (MKS3/TMEM67) cause nephronophthisis with liver fibrosis (NPHP11)” concluded that non-sense mutations in TMEM67 were responsible for causing the more sever diseases such as MKS and JS and missense mutations cause the milder NPHP. Applying this to the PKD lamb would mean it could be classed as having NPHP, however the phenotype we see in these lambs is closer to MKS. A possible explanation is that the presence of two missense mutations is causing the more severe phenotype.
Figure 4.2. Table showing the four diseases linked with the TMEM67/MKS3 gene and the phenotypes associated with them.

4.4. Characterisation of the disease phenotype.

Johnson 2005 gave a very thorough description of the disease phenotype found in flock A. Although the PKD flock was derived from flock B not flock A the characteristics of the affected lambs were found to be very similar. Importantly, when the phenotypes associated with defects of the TMEM67 gene are compared to that of the PKD sheep, a large number of similarities are apparent. The most notable characteristic in both the PKD sheep and reported human cases involving this gene are the dysplastic kidneys and the hepatic developmental defects.

An interesting note is that mutations in this gene often result in brain abnormalities ranging from occipital encephalocele to malformations of the brain stem. As this was known at the time of the 2009 dissections, the heads of the lambs were removed to allow for MRI scans. These scans would hopefully identify any central nervous system abnormalities, giving a complete picture of the extent of the disease. When an affected lamb survived birth it was noted that they seem disorientated in comparison to other lambs which would make sense if there were CNS malformations. However given the small number of affected lambs seen alive this cannot be taken as evidence of a defect at this time.
4.5. Other models of this disease

To date there are two model organisms for defects in the TMEM67 gene, the wpk rat and the bpck mouse.

The wpk rat is the most well known of these organisms. It was first described in 2004 by Gattone et al., (2004) as a model organism for polycystic kidney disease. It was described as developing proximal tubule and collecting duct cysts in utero, these cysts were seen to eventually dominate the kidney. Several extrarenal characteristics were also noted. These included hypoplasia of the thymus and spleen and CNS malformations such as agenesis of the corpus callosum and hydrocephalus. In 2006 Smith et al refined the region to 0.6 cM or a physical region of 2.05Mb. Thirteen conserved genes were identified in this region, upon sequencing of the novel gene LOC313067 they found a substitution, P394L, in exon 12. From this they identified LOC313067 as the wpk gene. In noticing the phenotypic similarities between the wpk rat and MKS patients the authors sequenced the human ortholog of wpk, TMEM67, in several MKS families. Upon identifying correctly segregating mutations in these families they came to the conclusion that TMEM67 was the causative gene in MKS3 individuals and in the wpk rat.

The bpck mouse was initially described by Cook et al., (2009). They described a mouse model with polycystic kidney disease, hydrocephalus, and morbidity by 3 weeks. They identified the causative mutation to be a spontaneous 295-kb deletion that includes the TMEM67 gene.

The deletion was mapped to a 0.6cM interval on chromosome 4 between the markers D4Mit261 and D4Mit19. This deletion encompasses six candidate genes. These are: cadherin 17 (Cdh17), protein phosphatase 2C magnesium dependent catalytic subunit (Ppm2c), an unidentified gene predicted from adult testis cDNA, Riken cDNA 1700123M08, Tmem67, and Riken cDNA C430048L16.

Looking at the candidate genes, TMEM67 stood out to the authors given the phenotypic similarities between the bpck mouse and MKS3 patients. In order to test the hypothesis that TMEM67 was the causative gene they performed transgenic rescue by using four overlapping BAC’s covering the deleted region. Inheriting the BAC’s containing TMEM67 rescued the mice and allowed for survival past a year with no signs of PKD, hydrocephalus, or elongated cilia.
Below is a comparison of the phenotypes of the wpk rat, bpck mouse, the PKD sheep, and human MKS3 patients.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Human MKS</th>
<th>Wpk rat</th>
<th>Bpck mouse</th>
<th>MKS sheep</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNS malformations</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>?</td>
</tr>
<tr>
<td>Cystic kidneys</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Gonad malformations</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hepatic dysfunction</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Metal retardation</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>?</td>
</tr>
<tr>
<td>Polydactyly</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Situs inversus</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Elongated cilia</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Biliary dysgenesis</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Retinal dysfunctions</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>?</td>
</tr>
</tbody>
</table>

Table 4.1. This table displays the phenotypes that have thus far been associated with defects of the TMEM67 gene in each animal thus far studied. The question marks in the sheep category indicate areas where insufficient data was available about the disease characteristic.

Of note is that none of the animal models of MKS3 defects have been observed to develop polydactyly. This could be explained by the findings of Consugar and colleges who concluded that polydactyly is more commonly associated with Meckel-Gruber syndrome when the causative gene is MKS1 and very rarely MKS3.
4.6. Other sheep models of disease

Over the years many ruminant models of disease have been developed for a wide range of conditions. A large number of conditions affecting bovine flocks have been proposed as models for human diseases, however in a 2008 editorial Agerholm suggested that sheep may prove to be a more useful model animal. He drew this conclusion based on sheep being easier to handle and care for, being less expensive than cows, and they have a shorter gestation and a higher average litter size which gives more progeny in a shorter amount of time. Although the advantages of sheep over cattle as disease models are clear, they would still need some advantages over rodent models to overcome the differences in price and ease of management. One advantage sheep have over rodent models is their longevity which allows for studies of chronic disease to be carried out. Another benefit comes from the size of sheep in relation to humans. In a 2010 paper by Meeusen et al the use of sheep as models of human asthma and other respiratory diseases was discussed. The authors pointed out that due to the similarities in size between humans and sheep the same equipment and procedures as used in humans is able to be used in sheep, thus facilitating translation of findings into the clinic. The Meeusen paper also includes a list of eight sheep models of respiratory disease. Although it is not a complete list it demonstrates that sheep are already a widely used and useful disease model. Additionally much of what we currently know about lung physiology under both normal and pathological conditions was derived from studies conducted in sheep which have lungs that share many characteristics with human lungs. As well as respiratory diseases there are also a large number of sheep models of neurological diseases. Sheep are useful for studying neurological diseases for much of the same reasons as they are useful for respiratory diseases. Their larger size makes surgical manipulation of their nervous system easier than it would be in mice. An example of this is von Koch et al., 2005 in which they surgically induced a lesion in foetal sheep at 75 days gestation in order to create a sheep model of myelomeningocele, which is the incomplete forming of the neural tube resulting in malformations of the spine. This study identified the sheep model as being useful in developing techniques to deal with myelomeningocele related complications such as leakage of cerebrospinal fluid.

From these examples we see that sheep are a useful model organism for human disease, however more work is required to identify potential candidate diseases in sheep.
4.7. Conclusions

Homozygosity mapping using the SNP50 ovine BeadChip was essential for identifying putative, causative genes in a lamb disease model displaying multiple system defects, most notably polycystic kidneys with hepatic involvement. Mutations in the TMEM67/MKS3 gene were identified and the hypothesis that they are causative was unable to be falsified by bioinformatic analysis. The disease characteristics of the sheep are similar to the symptoms seen in mutations of this gene in humans and other animal models.

In conclusion the evidence supports the hypothesis that the mutations I638N and I644S in the TMEM67 gene are causative in this large animal model of PKD with multiple extra renal symptoms. TMEM67 or meckelin has been linked to several ciliopathic syndromes, making this a potentially very useful large animal model for studying ciliopathies.

4.8. Future directions

Future prospects for study of the PKD sheep include more work to characterise the disease phenotype. This would involve brain scans of affected animals to look for CNS malformations as these are found in all other species displaying this disease. An examination of the retina of the sheep may also provide some insights as the retina of the wpk rat are found to be defective. Examining weather the sheep show signs of mental deficiency may prove difficult given that very few are found alive and measuring the mental acumen of sheep would likely be very imprecise, if it is possible at all.

Given that the mutations identified by this work are yet to be confirmed as causative it would be useful to perform in vivo site directed mutagenesis. This could be done in a ciliated cell line from any animal containing the TMEM67 gene as this gene is seen to be highly conserved the results should be transferrable. Using this it would be possible to confirm one, or both of the sequence variants as causative if a phenotype of interest was observed. Such a phenotype might be the elongation of cilia, or alternate localisation of the Meckelin protein in the mutants compared to the wild types. Additionally sequencing of exons 9 and 10 would provide more evidence for I638N and I644S being causative, assuming no sequence variant was found in either of those exons.
References


### Appendix A: Reagents used

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Cat#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarose powder (Cambrex)</td>
<td>50004</td>
</tr>
<tr>
<td>Bromophenol blue (Sigma)</td>
<td>B8026</td>
</tr>
<tr>
<td>DNA ladder, 1kb plus (invitrogen)</td>
<td>10787-018</td>
</tr>
<tr>
<td>dNTPs (Invitrogen)</td>
<td>10297-018</td>
</tr>
<tr>
<td>Ethanol (Shalau)</td>
<td>ET0016</td>
</tr>
<tr>
<td>Ethidium bromide (Invitrogen)</td>
<td>15585-011</td>
</tr>
<tr>
<td>Ethylenediaminetetraacetic acid (EDTA) (BDH)</td>
<td>100935V</td>
</tr>
<tr>
<td>Glacial acetic acid (BDH)</td>
<td>100015</td>
</tr>
<tr>
<td>Isopropanol (BDH)</td>
<td>102246L</td>
</tr>
<tr>
<td>Magnesium Chloride (Invitrogen)</td>
<td>10966-034</td>
</tr>
<tr>
<td>PCR buffer, 10x (Invitrogen)</td>
<td>10966-034</td>
</tr>
<tr>
<td>Platinum Taq (Invitrogen)</td>
<td>10966-034</td>
</tr>
<tr>
<td>Primers (Invitrogen)</td>
<td>N/A</td>
</tr>
<tr>
<td>Proteinase K (Invitrogen)</td>
<td>AM2564</td>
</tr>
<tr>
<td>Tris (Invitrogen)</td>
<td>15504020</td>
</tr>
<tr>
<td>Xylene cyanole FF (Sigma)</td>
<td>X4126</td>
</tr>
</tbody>
</table>

### Kits

<table>
<thead>
<tr>
<th>Kits</th>
<th>Cat #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promega wizard DNA isolation kit</td>
<td>A1120</td>
</tr>
<tr>
<td>Nuclei lysis solution</td>
<td></td>
</tr>
<tr>
<td>Protein precipitation solution</td>
<td></td>
</tr>
<tr>
<td>DNA rehydration solution</td>
<td></td>
</tr>
<tr>
<td>RNase solution</td>
<td></td>
</tr>
</tbody>
</table>
Qigen QIAamp blood midi/maxi kit 51185
   QIAamp midi spin columns
   Collection tubes
   Buffer AL
   Buffer AW1
   Buffer AW2
   Buffer AE
   QIAGEN protease

QIAquick gel extraction kit 28706
   Qiagen QIAquick spin columns
   Elution buffer (EB)
   Wash buffer (PE)
   Binding buffer (PB)

Appendix B: Formulations and solutions

10x TAE
   96.8g Tris
   22.8ml Glacial acetic acid
   14.90 EDTA
   dH2O to 2L

1x TAE
   50μL Ethidium bromide
   200ml 10x TAE
   dH2O to 2L

Loading dye
   4g Sucrose
20mg (0.2%) Bromophenol Blue
20mg (0.2%) Xylene cyanole
dH2O to 10ml

Agarose gel
3g agarose powder
200ml 1x TAE

10% Neutral buffered formalin
10ml formalin
90ml dH2O
Sodium phosphate to pH of 7.0-7.2
Appendix C: Pedigrees

Pedigree Legend

Unaffected male

Carrier male

Affected Male

Unaffected female

Carrier female

Affected female

Unknown sex
2005 born lambs from ram 98-8587

- Pink36
- Green 3
- P7015
- Pink32
- 96-642
- Pink22
- 01-17
- 01-1
- 98-8587
- 01-12
- 98-8587
- 01-14
- 98-8587
- 01-19
- 98-8587
- S3
- 8001
- 05/21
- 05/22
- 181/35
- 05/34
- 05/19
- 05/20
- 05/25
- 05/26
- 05/36
- 05/37
- 05/38
- 05/29
- 189/190
2005 born lambs from ram 01-2

S10
Pink 2
01-32
01-26
S4
S14
S11
01-2
S5

06/1
05/2
05/24
no name
no name
189/3
05/4
06/5
05/6
184/185
06/40
05/11
05/12
05/23
05/27
05/28
05/33
2007 born lambs from 6005

3001
01-26
01-33
01-43
01-34
01-40
01-41
6005
Scott 4

07-7
196
07-6
07-5
07-8
479
498
07-13
278
07-13
07-11
82
07-2
2008 born lambs from ram 6026

22
27
33
41
Blue kenmore 200
14

2008 born lambs from ram 6093

6020
6093
6017
31
16
2009 born lambs from ran 6005

- 01-23
- 07-6
- 5011
- 01-40
- 01-3
- 6005
- 01-26

- 09/H
- Not tagged
- 09/F
- Not tagged
- 09/G
- Not tagged
- 09/D
- 09/E
- 09/A
- 09/B
- 09/C
- Not tagged
Appendix D: Genes found in homozygous loci

Chromosome 4
3-hydroxyisobutyrate dehydrogenase
Even Skipped homeobox 1

Chromosome 9
RNA binding motif protein 12B
Transmembrane protein 67 (TMEM67/MKS3)

Chromosome 11
Dishevelled, dsh homolog 2
PHD finger protein 23 (PHF23)
GABA(A) receptor associated protein (GABARAP)
Dullard homolog (DULLARD)
Claudin 7 (CLDN7)
Solute carrier family 2, member 4 (SLC2A4)
Y box binding protein 2 (YBX2)
Translation initiation factor 5A-like1 (EIF5AL1)
Translation initiation factor 5A (EIF5A)
G protein pathway suppressor 2 (GPS2)
KIAA1787 protein (KIA1787)
Centaurin beta 1
Potassium channel tetramerisation domain containing 11 (KCTD11)
Transmembrane protein 95 (TMEM95)
Tyrosine kinase, non-receptor 1 (TNK1)
Phospholipid scramblase 3 (PLSCR3)
Neuroligin 2 (NLGN2)
Transmembrane protein 102 (TMEM102)
Fibroblast growth factor 11 (FGF11)
Cholinergic receptor, nicotinic, beta 1 (Muscle) (CHRN1)
Zinc finger and BTB domain containing 4 (ZBTB4)
Acyl-malonul condensing enzyme 1 (AMAC1)
Polymerase (RNA) 2 (DNA directed) polypeptide A
Tumour necrosis factor (Ligand) superfamily member 13
TNFSF12-TNFSF13
SUMO/Sentrin/SMT3 specific peptidase 3 (SEMP3)
Translation initiation factor 4A
CD68 molecule
Mannose-P-dolichol utilization defect 1 (MPDU1)
SRY (sex determining region Y) – box 15 (SOX15)
Fragile X mental retardation autosomal homolog 2
Spermidine/Spermine N1-acetyltransferase family member 2 (SAT2)
Sex hormone binding globulin (SHBG)
ATPase NA/K transporting, beta 2 polypeptide (ATP1B2)
Tumour protein P53
WRAP53
WD repeat domain 79
Ephrin-B3 (EFN4)
Dynein, axonemal, heavy chain 2 (DNHA2)
Jumonji domain containing 3, histone lysine demethylase (JMJD3)
Chromodomain helicase DNA binding protein 3 (CHD3)
Potassium voltage-gated channel, shaker-related subfamily, beta ember 3 (KCNAB3)
Centrobin, centrosomal BRCA2 interacting protein (CNTROB)
Guanylate cyclise 2D, membrane (retina specific)
Arachidonate 12-lipoxygenase, 12R type (ALOX12B)
Arachidonate lipoxygenase 3 (ALOXE3)
Hairy and enhancer of split 7 (HES7)
Period homolog 1 (PER1)
Vesicle associated membrane protein 2 (Synaptobrevin)
Transmembrane protein 107 (TMEM107)
Aurora kinase B (AURKB)
Phophoribosylformglycinamidin synthase (PFAS)
RAN guanine nucleotide release factor (RANGRF)
Solute carrier family 25, member 35 (SLC25A35)
Rho guanine nucleotide exchange factor (GEF) 15 (ARHGEF15)
KRAB-A domain containing 2 (KRABA2)
Ribosomal protein L26 (RPL26)
nudE nuclear distribution gene E homolog (A-nidulans)- like 1 (NDEL1)
Myosin, heavy chain 10, non-muscle (MYH10)

Appendix E: PCR gel pictures

<table>
<thead>
<tr>
<th>Lane</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product</td>
<td>Ladder</td>
<td>R</td>
<td>Q</td>
<td>L</td>
<td>K</td>
<td>J</td>
<td>43</td>
<td>41</td>
<td>39</td>
<td>34</td>
<td>H</td>
<td>33</td>
<td>G</td>
<td>28</td>
<td>26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lane</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product</td>
<td>Ladder</td>
<td>M</td>
<td>F</td>
<td>E</td>
<td>D</td>
<td>O</td>
<td>13</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>P</td>
</tr>
</tbody>
</table>

PCR for exon 2 was carried out by Dr Cherie Stayner
Appendix F: TMEM67 Sequence

Exon 1 (“R”)

GTGGCGATGGCCGCTCGGTCACCTCTGTGCAGATGACTGTCGCTGGCCGCTCGGCTCGGGCAGCCAGAGACGTGCGGCCACAACCAGTACTTCGATATCTCGCGCTGTCTCCTGTGTGGCGTGTGGGCCAATCAGAGGCAGGATGCCCGA

Exon 2

CAKGKGTTCGKGGGACATTTTTTACAAATATATCAGCCTCYYCCATTATTTAATCA

Exon 3 (Primer set “Q”)

AAAGGTGTACTGAAGATGGCTGGAACTGTATTTCTTGCCCTGGTTATATACTGCAGAAGGAAAATGCCACTGTCCCACTGGCCATATTTTAG
Exon 4 (Primer set “L”)

TGGAAAGAAATGTTATGGAAAACATTGGTTGTCTCAAGCAACCTTGTGAGCTCT
GTGATGAAAGTGAACACCTTTTTCACAGTAGCAAATGCTTTAGGAAACAG

Exon 5 (primer set “K”)

GTGTGCCTCGATGTGAACCAACATTCATTAATACCCAGCAGGCCCTGCTGATG
TTCAAGAACCTAAACATTTTA

Exon 6 (primer set “J”)

ACAGGGGGCTTTGTGTTTCAGCAGCACAGGGAATTTCCTTACGTATGATT
TCAGCTGCACGGTTATGGAGA
Exon 7 (Primer set “43”)

CCAACAAGCAGCTGCTGATGACTGCAAATACCTTTGCAAACCATGCTGAAGT
GAAAGATATGCCC

Exon 8 (primer set “41”)

CAAAAGGAAACAGAATGAACAGTGCCCAGTGACAGTGATTTTCAAAAGAT
AACCCTGAAACAATCCACATGCATCATAAAAGTAGAGGAGTATAAAATTCAT
GTTCATTACACACATATTTCACACAGCCTTGGAAGATGTTAGATTGCCATAC
AC

Exon 11 (primer set “39”)

CTTTGTCCAGACACAGAGCGAGACGACTAAATGTGCTGCTATTCCCTTTTGAACA
ACTTATCAACAAA
Exon 13? (primer set “H”)

ACAGTAGTTCGGGCAAGTGGCTTCTGACTCGGCGCATTTTCTTAGTGGATA
CACTAAGTGGAAGGGAAAAATGACTTAGGAGAGCCAGCGAAATATTAGAG
GTCGCCACCCAAATATTCACCTGAG

Exon 12? (primer set “34”) Resequence

TTTCATCGGTGTATTCAGGTGCACATCATAAGGTACTGAGTGAGGAAAAAGC
CGGATAAGGTCAGGATGGGAATCTCACA

Exon 14 (primer set “33”)

TCCACCTTGTCACCCACACAAATGGAACACATTTACCTCCCTTATAATAC
CATTTGTCTACAGTGACGTTGATATCAAGATCCCAACACGCCAGTCGTGAA
G
Exon 15 (primer set “G”)

GTTTCTTTTCTCAGTCACATACGAAATGGATCAAAGAGAAGCACAAGTCCAG
ACAGAT

Exon 16 (primer set “28”)

ATTGCTTTTAGGTGTGTTTGCGTGACTAGCTGTTTTATCATCTCTTTTGAAG
ACAGCAGGTTGGAAGAGACGCATTGGGAGTCCCATGATTGATTTACA

Exon 17 (primer set “26”)

ACAGTTACGAAATTTCTTGGTTGACTATGCTGGTGGATCTGCGCAATGTTTTC
TTTATCATCACTGAGAAACAGGTCTTTACTGGCTTTTTTCTTTAAA
Exon 18 (primer set “rM” and “f24”)

CTTGCCCTAAAGGACTCTCTTTCCCAAGGGTTCCAAAAGCCTTTAACTTCTTTATAAGGAAAGGTATCTCCAAACTGTAATTGGTTTGTCC

Exon 19 (primer set “F”)

GCTCTGCAATTTTTGCATAAGCTCATCTCCAGATTACCATAGACATATTCTTTATCGAT TGGGAGCGGCCTAAAGGGAAGGTTCTTAAAGCTGTTGAAG

Exon 20 (primer set “E”) putative mutation

GTGAGGGTGGTGTCGCCGGAAGTGCTACTGTTCCTGTAAGCATATGGAGAATTATTGTCATGAGAAAATCTAAACCCTTTCAAGTACTTACTGTCCTCTTTTGGAG
Exon 21 (primer set “D”)

GTTGTGGGATTTCAAGAAGTATAGATGGACTCATCTCTCTAGTCTTTCCAGAAGCCCACCTAGCTACATAGCTCCTTACAGCCGCATTTTGAGATATGCGGTGTCTTCTGCTCTTTGGCTGGTCATTGGAATTATACAG

Exon 22 (primer set “O”)

GTGTTCCTTTGGGTGTCCTCTATTGAGAGATTTATAGAATAAATTCGACAGTTTGTTGATTTGTGCTGTATGAGTAAT

Exon 23 (primer set “13”)

ATATCAGTTCTTCTGTTATCCACAGATGTTTGGCTACTACATCCATGGTAGATCACTACATCCCATGGTAGATCCAGTACATGGGCATGGTGATACTAATATGGGAAGAATGAATA
Exon 24 (primer set “C”)

GAAAATTTATGTAGCCAAAAGAGGTTTGTGCGCAATACAGATGTCAGACT
TTTCAGATTGCAATTTTCTAGCCAGATGAGGCAACACTATGACAGAATTCAT
GAGACACTAACAAGG

Exon 25 (primer set “C”)

AAAAATGGCCCTGCGAGACTCTTGAGTTCCTCAGGAAGTACTTTTGAGCA
GAGTGTAAAAGCATATCATACTATGAATAAATTTCTTGGCTCTTTCATCGAT
CAT
Exon 26 (primer set “B”)

GTTCATAAGGAAATGGACTACTTTTATAAAGATAAAATTGCTTCTTGAAAGAA
TTCTTGGAATGGAATTCATGGAACCAATGGAAAAAGCATCTTTTACATAG

Exon 27 (primer set “A”)

ATGAAGGATATTCTTTCAGCAATGTTCTGTATTATGGAAATGAAGCTACTCT
TCTTATTTATGATCTGATGTTCTTCTGTGTTGTGGATTTGGCTTGCCAAAAT
TTTGTTTTAGCAGCCTTCCTTACACATCTACAAACAG

Exon 28 (primer set “P”)

111
ATTTTTAGATTTATTCGAATGACAGACAGAAGATTTGGCATCCAAAA
CACTTGGTGATCCAAAAATTTCTGATTTAAAAATGATGATAGATAAGGCAAT
TATAATCGTGCAAAAAAAGGTCATGATTATCAAGTCAGTTTCTATATT
TTTTAAAACCTGTAATACAAATATCTATATTTTTTATTTTGTCTACA
GGAARATTATTTTTTTAATCTACAAGAATATG
TAATATGATAAATAATGAACAATATTGTTTTCCTGCATTTGAGTAGATAATTC
ACCAGCTCATAAACACTARATACGTCAACAAKGGGACCAAAAA

Ovine Meckelin protein sequence

M A A R S T L S A E S V T L F L L L V L P R V S Q A T F L F P F R Q P E T C G H N
Q Y F D I S A L S C V A C G A N Q R Q D A R T S C V C L P G F Q M I S N N G G P A
E R N V N G T L S Q A T C E L C D E S N F T V A N A L G N C V R C E P T F I
N T S R S C A C S E P N I L T G L C F S T T G N F P L R M I S A A R Y G G I S F T S
A W F A K Y L Q S S A A C W V Y G N L T S C Q A V G N M C V M N M N Y D S
S T F D A C G L F Q L S K I L L H W A L F I L F X F R Q N L P W L F Y G D Q L G L
A P Q I L S T T P L N F S L K G N T N T K L K F V A A Y D V R G N F L K W Q
T L E G G I L C P D T E T R L N A Y S F G T Y Q S S S G K W L L T R R I F
S V K V S F S V T Y E M D Q R E A Q V Q T D I A L G V L G L A V L S L L K T A
W L I F F K A Q K S V S V L L P M I Q E E R F V T Y V G A C A L F C V C H S Q
P V S I W R T Y F V A N E W E I Q T V R K I N P L F Q V L T V L F L L E V V G F K
F F G V F Y E R F E D K I R Q F V D L C C M S N I S V L L S H R C F G Y Y I H G
K F L G S F I D H I V K D A F H W F H F H F H S K N S F K K Q F I F Y K V H F L
M N E G Y S F S N V L Y Y G N E A T L L Y D L M F C V V D L A C Q N F V L A A
F L T Y L Q Q E I R F R I R N A V G Q K N L A S K T L V D Q R F L I

112